

LONDON  
SCHOOL of  
HYGIENE  
& TROPICAL  
MEDICINE



LSHTM Research Online

Douglas, Hannah-Rose; (2003) Economic evaluation of complex multidimensional health services : the case of palliative day care. PhD thesis, London School of Hygiene & Tropical Medicine. DOI: <https://doi.org/10.17037/PUBS.01379945>

Downloaded from: <https://researchonline.lshtm.ac.uk/id/eprint/1379945/>

DOI: <https://doi.org/10.17037/PUBS.01379945>

**Usage Guidelines:**

Please refer to usage guidelines at <https://researchonline.lshtm.ac.uk/policies.html> or alternatively contact [researchonline@lshtm.ac.uk](mailto:researchonline@lshtm.ac.uk).

Available under license. To note, 3rd party material is not necessarily covered under this license: <http://creativecommons.org/licenses/by-nc-nd/3.0/>

<https://researchonline.lshtm.ac.uk>

**LONDON SCHOOL OF HYGIENE AND TROPICAL MEDICINE**  
**University of London**

**Economic evaluation of complex, multidimensional health services:  
the case of palliative day care**

*Thesis submitted for the degree of PhD*

**Hannah Rose Douglas**  
**Health Services Research Unit**

## **Abstract**

Palliative day care is offered to patients with a life-limiting illness. It provides a range of services to meet the varying needs of individuals. Consequently, the service is multidimensional and the outcomes are hard to define. This means this service operates at a level of complexity that is different from other health care services. This presents a challenge for evaluation since the activities differ across individuals and change over time, depending on their needs. This challenge is not unique to palliative care and is relevant to the evaluation of other services. The published guidelines for health economic evaluation have not explicitly taken into account the specific issues relevant to evaluating complex services. It is argued in this thesis that the 'health gain' approach is problematic and a preference-based approach may reveal more useful evidence for policy-makers. This has not been fully considered in empirical studies.

A study was undertaken using health-gain and preference based approaches to evaluate a new palliative day care service. The EQ-5D health-related quality of life instrument was used to detect differences in outcome between patients who attended a palliative care day centre and those who did not. The instrument did not detect significant differences over time. A choice experiment was also undertaken in four PDC centres. This approach estimates the relative preferences that respondents expressed for specific service attributes. The attributes chosen for this study were: opening hours, access, specialist therapies; medical support; hairdressing and bathing. The results showed that specialist therapies were relatively more valued, and that hairdressing and bathing were not important in decision-making. Access to medical care was less important than access to specialist therapies.

The thesis critiques health gain approaches for services that have a broad range of hard-to-define goals and aim to meet individual needs. The choice experiment provided insights into how services are valued where these insights cannot be derived from other economic evaluation approaches. Research methods that reflect people's preferences may provide important analysis where very little evidence has previously existed.

## **Acknowledgements**

First and foremost, I would like to thank my supervisors Charles Normand and Pauline Allen for the support they have provided in helping me to develop my interest in health economics and policy, and in the specific ideas that underpin this work. Also, thank you to Jenny Roberts for reading the entire thesis and for valuable insights and comments. Also, I would like to acknowledge the help of Jan van der Meulen who provided statistical support and advice in the design and analysis of the choice experiment.

This thesis is part of a collaborative project to evaluate palliative day care in the UK. I would therefore like to thank Irene Higginson and Danielle Goodwin at the Department of Palliative Care and Policy at Guys, Kings and St Thomas' Medical School for their support and teamwork throughout the project. I would also like to acknowledge the support of the patients, staff and managers of the palliative day care centres where I carried out this work.

I have also had a great deal of support from friends outside the London School of Hygiene and Tropical Medicine who have encouraged my academic career and interests, including Tessa Adams, Myra Barrs and Jonathan Elford. Emily Steel also undertook the additional task of reading through the thesis at the end, for which I am very grateful.

My parents, Stuart and Carolyn Douglas, my sister and brother, Jessica and Ben, and all my family have encouraged me from the beginning, and I would like to thank them all for their loving support.

Finally, I would like to recognise the enormous contribution to the completion of this work of Mark Atherton, whose love and support has given me the courage and confidence to see this through to the end.

The thesis is dedicated to the memory of my mother.



## Table of contents

	Page
Abstract	2
Acknowledgments	3
Preface	7
Chapter 1: Introduction	9
Chapter 2: Theory and application of welfare economics and economic evaluation	27
Chapter 3: Review of the economic literature in palliative care research	76
Chapter 4: Empirical investigation of the outcomes of complex services: methodological issues	104
Chapter 5: Methods of empirical research into palliative day care	154
Chapter 6: Results of empirical research into palliative day care	178
Chapter 7: Discussion of the empirical findings	201
Chapter 8: Conclusion	221
Bibliography	243
Appendix A Summary of research evidence from a systematic review of economic studies in palliative care	273
Appendix B Copies of patient information sheets and consent forms	275
Appendix C Research instruments used in the clinical effectiveness and EQ-5D study	279
Appendix D Research instrument used in the choice experiment study	288
Appendix E STATA and SPSS output for the probit models	299

**List of tables****Page**

Table 3.1	Characteristics of the five palliative day care centres in the North Thames PDC study	88
Table 4.1	Values for estimating weighting for each state of health using values estimates for the UK general population	111
Table 5.1	Breakdown of activities and structure of the day by PDC centre	165
Table 5.2	Final attributes of PDC included in the choice experiment model	170
Table 5.3.	Specification of attributes and levels for the probit model	174
Table 6.1	Completed patient interviews in the EQ-5D study	180
Table 6.2	Numbers and proportions of respondents with completed interviews for each data collection stage	179
Table 6.3	Patient socio-demographics (completed 1st interview)	181
Table 6.4	Cost of four weeks' care: patients with only the first set of data at baseline	182
Table 6.5:	Cost of four weeks' care: patients with data for interview 1 and 2	182
Table 6.6:	Cost of four weeks' care: patients with data for interview 1, 2 and 3	182
Table 6.7	P values from Analysis of Variance (ANOVA) tests of difference in mean costs of care at time 1,2 and 3, by palliative care group.	183
Table 6.8	Percentage of patients at level of each domain of EQ-5D, by intervention group, at interview 1, 2, and 3, compared with the general UK data	186
Table 6.9	Pearson correlation coefficient scores and p-values for the 'Pain' domain on POS and EQ-5D, by palliative care group and by interview	187
Table 6.10	Pearson correlation coefficient scores and p-values for the 'anxiety/depression' domain on POS and EQ-5D, by palliative care group and by interview	187
Table 6.11	Mean quality of life score at interview 1, and results of the t-test, for patients completing only interview	189
Table 6.12	Mean quality of life score at interview 1 and 2 and results of the t-test, for patients completing 2 interviews or more	189
Table 6.13	Mean quality of life scores for patients who completed all three interviews	189
Table 6.14	Respondents in the PDC group who said they had 'changed' as a result of palliative day care: proportion of responses at each interview	192

Table 6.15	The proportion of all patients' interviewed for the choice experiment. by centre	194
Table 6.16	The results of the random effects probit model (without the two hold outs): All patients including those manifesting dominant preferences	195
Table 6.17	Dominant preferences observed in the data, all patients	197
Table 6.18	The results of the random effects probit model: Only those patients who did not display dominant attributes	197
Table 6.19	Activities accessed in PDC by respondents in the study	198
Table 6.20	Probit model interaction terms explored in the data, all patients scenario 1-8 only	199

#### **List of figures**

Figure 1	Mean costs for patients in each group over the three time periods	184
Figure 2	Means costs for patients in each group over the three time periods	184
Figure 3	Box plot of EQ-5D composite scores for patients who had data for one interview only	190
Figure 4	Box plot of EQ-5D composite scores for patients who have data for at least two interviews	190
Figure 5	Box plot of EQ-5D composite scores for patients who have data for at three time points	191
Figure 6	Patient reports of most important thing: frequency of themes at each interview	193

## Preface

The evidence presented in this thesis is based on a multi-disciplinary palliative day care study undertaken by a team of researchers at the London School of Hygiene and Tropical Medicine (LSHTM) and the Department of Palliative Care, Guy's King's and St Thomas' Medical School (GKT). The study team members were: Charles Normand (Principal Investigator, LSHTM), Hannah-Rose Douglas (LSHTM), Irene Higginson (Principal Investigator, GKT), and Danielle Goodwin (GKT). The funding for the study was obtained from South East Regional NHS Executive, R&D Project reference SEO 006.

Each of the team members undertook various roles. HRD and DG managed the study on a day to day basis and liaised with the study sites. DG's main focus was the clinical effectiveness study, HRD's focus was on the economic evaluation. The consent form, patient information sheet were designed by DG with contribution by HRD. Ethics Committee approval for the study was obtained by IH on behalf of the team. IH was the main liaison with the funders. HRD was responsible for collecting data for the cost analysis and for the choice experiment. This involved designing and implementing the pilot study and main experiment. Additional statistical support was provided by Jan van der Meulen at LSHTM. Interviews with patients were undertaken by trained interviewers with palliative care nursing experience. The choice experiment interviews at one palliative day care centre (out of four) were undertaken by HRD. EQ-5D data and patient resource use data were analysed by HRD. The choice experiment data and cost data were analysed by HRD. The palliative care outcome data and qualitative data were analysed by DG.

Some of the qualitative data analysed by DG is presented in this thesis for completeness. This work has been fully attributed to DG in the main text and footnotes.

The project had two main aims: to explore the clinical effectiveness of palliative day care and to consider issues of cost-effectiveness. The clinical effectiveness data used published and validated palliative care outcome scales. The cost-effectiveness analysis used the EQ-5D generic health-related quality of life instrument to explore whether it would be possible to obtain measures of effectiveness in units of quality-adjusted life years. The expectation was that this methodology would not yield useful results for a group of patients who are generally well enough physically to attend a palliative day care unit, and may only have a few weeks or months to live.

A previous study undertaken by the same research teams in LSHTM and GKT from 1998-2000 (The North Thames study, referred to in this thesis) had both a qualitative and quantitative component. The quantitative study found that there was no difference in health-related quality of life using published palliative care outcome instruments. Similarly, there was no significant difference in costs between those who attended day care, and those who did not. It was postulated that the EQ-5D study design and palliative care outcome instruments would not show important differences in outcome between palliative care groups.

The other focus of the empirical research undertaken by HRD in this second palliative day care study was to explore preference-based, multi-attribute methods of economic evaluation. The idea for using this approach came from experience of HRD in undertaking in-depth observational and interview data collection as part of the developmental work for the first North Thames study. The strong preferences for palliative day care had been expressed by patients in the qualitative study. Respondents had also indicated that they might be willing to pay for some aspects of palliative day care, such as hairdressing. This led HRD to consider whether a preference-based approach in the economic evaluation would yield insightful findings. The choice experiment methods in palliative care research presented in this thesis is the result of a process of exploring the problems with a cost-consequence study (as undertaken in the North Thames study), the EQ-5D approach and other preference-based approaches to evaluating complex services .

The choice experiment was planned to be the main component of the economic evaluation in the second study, on which this thesis is based. Palliative care outcome instruments were included in the clinical effectiveness component of the study since the second study used a more robust study design. The EQ-5D approach was rejected for the economic evaluation in the first instance since the experience from the North Thames study suggested that it would not be sensitive enough to detect the subtle changes in health-related quality of life appropriate to palliative day care. However, the research team were challenged by the funders of the study to provide some evidence that the EQ-5D would not be a useful instrument in this context. It was proposed by them that the EQ-5D instrument should be included in the study since the effort of collecting and analysing data for EQ-5D (a five-item questionnaire) alongside the palliative care outcome instrument was marginal. The results of the EQ-5D study are presented in this thesis as they complete the story of the investigation of palliative day care and help to explain the development of ideas that led to the choice experiment approach.

## Chapter 1

## Introduction

### Scope of the thesis

This thesis is about the economic evaluation of health care interventions that are complex, multidimensional and have subtle and varied outcomes that are hard to define or measure. In particular, it considers the contribution of welfare economics and other economic paradigms to the understanding of how to undertake evaluation of complex interventions in health and social care. An empirical study of palliative day care that incorporates different research methodologies representing different ways of conceptualising value will be used to explore these issues.

The main argument that is put forward in this thesis is that for some areas of health and social care, economic evaluation is more difficult than in others and that this relates to the difficulty of conceptualising and quantifying the outcomes of complex services. There are particular characteristics of these services and how they are accessed by patients that distinguish them from mainstream services. As a consequence, straightforward approaches to evaluation may not provide useful information for comparison across different kinds of care and for decision-making about resource allocation.

Like all research, economic evaluation can be difficult to do well, but in the evaluation of complex services, there are particular challenges. The nature of this complexity and the nature of the problems of evaluation in these contexts is the subject of the thesis. The main argument that is explored is whether approaches that are based on the theories of choice, trade and the satisfaction of preferences represent a way forward in evaluation.

### SECTION 1. INTRODUCTION TO THE THEMES EXPLORED IN THE THESIS

The need to compare the cost and value policies that aim to improve or sustain welfare in different ways is fundamental to making choices about how resources are used. In a democratic society, decision-makers are charged with making choices that are may be seen as fair to society. They must balance the interest of future people against the interests of people currently alive, the importance of welfare in old age against welfare in youth, and of the value of

prolonging life against the value of improving the quality of life. Therefore the notion of valuing and comparing welfare across different types of policy and uses of resources is inherent in the process of decision-making.

Individuals make comparisons of welfare in their daily lives by choosing between current and future consumption and by weighing up the costs and benefits of different bundles of goods to maximise their own or their family's welfare. These forms of private consumption reflect the individual's values and tastes, and it is usually unnecessary for an individual to have to justify their choice of consumption to others. In nearly all situations, the satisfaction of individuals' preferences is seen as contributing to the good of society as a whole.

In the use of public resources, however, there is a need for more transparency in decision-making. Therefore choices between different uses of resources have to be based on explicit criteria that are rational and defensible, that is, on which reasonable people can agree. It is necessary to define what is good for individuals and for society, and how this good can be maximised within a given level of resources. These are problems of moral philosophy as well as of economics.

Welfare economics is a theoretical framework for considering social costs and benefits. It is based on specific notions of how individuals behave, what motivates them, and the meaning behind the concept of one thing being 'better' than another; of a change in policy being for the good, making society better off, or contributing more to society's overall welfare. Within this framework, individuals are assumed to be self-interested, knowledgeable, welfare-maximising consumers. Preference satisfaction is the fundamental criterion for deciding what is 'good' for society. Given a choice of A or B (two products, services, interventions or policies), A can be said to be 'better' than B if and only if individuals prefer A to B. The strength of preference for A over B can be measured by how much an individual, or the sum of all individuals, would pay to consume A rather than B, (or would need to be compensated to consume B rather than A). Therefore the individual is the locus of decision-making and preferences are expressed through the medium of money, since money is the most divisible, transferable form of wealth.

Furthermore, since welfare economics is based on the idea of satisfaction of individuals' preferences, the psychological processes that lead to A being preferred to B by a single individual are not known. Within this framework, it is not necessary to understand *why* the

individual prefers A to B, only to know that A is preferred. Welfare economics also lays out specific and explicit criteria to decide whether a change in circumstances (for example a new intervention, a change in policy, a new call on resources) improves overall social welfare. If the benefits of a change, measured in monetary terms, to some members of society outweigh the overall costs to other members of society, then the decision rule is that it ought to be adopted. Society is better off after such a change, regardless of who gains and who loses. The theory therefore provides a rationale for behaviour and criteria for making socially beneficial decisions that is internally coherent and can be defended its own terms.

There has been strong criticism of the underlying axioms of welfare economics and questions raised about the strict assumptions of 'rational hedonism' of individuals. It is argued that self-interested, welfare-maximisation has failed to explain important aspects of human behaviour in the real world, such as communitarianism, caring for others, or altruism. Also, the measurement of outcomes in monetary terms has been seen as unachievable in some contexts and even inherently unethical where issues of life and death are concerned. These tensions are played out in the debates about how economic evaluation ought to be undertaken and interpreted.

### **The economic evaluation of health and social care**

The economic evaluation of health and social care has developed in a welfare economics paradigm. Within this framework, public sector economists have had to address some general problems of finding ways to express the benefits of health and social care, and measuring and comparing different kinds of benefits across interventions that aim to improve welfare in contrasting ways. Monetary valuation would indicate what an individual or society might be willing to sacrifice (in terms of other consumption forgone) in order to access a particular service or intervention. However, monetary valuation of the benefits of health care has presented a serious methodological challenge to evaluators. Instead, health economists have identified other means of expressing the benefits of care, by using notions of improved length or quality of life to conceptualise the outcomes of different interventions. Therefore the idea of value or revealed preference, central to welfare economic analysis, has been less important in the economic evaluation of health care interventions. The value of the outcomes derived from health (and this could also be applied to social care) has usually been implied rather than explicitly argued.

For some types of interventions this has not been a major problem: the definition of the outcome or purpose of the intervention has been widely accepted. The welfare or benefits derived from



particular interventions are well-understood by professionals and the public, and the relationship between the intervention and the outcome is straightforward. The value of the benefits to one person of, for example, a hip replacement operation can be assumed by most reasonable people to be comparable to the value of the operation to another. The comparison of welfare between individuals would be considered to be relatively unproblematic in this context.

Important progress has been made in economic evaluation from assuming the overall comparability of welfare. This has led to the design and validation of generic measures of outcome that are comparable across individuals and across different types of interventions. Approaches such as the EQ-5D have been incorporated into economic evaluation to generate estimates of additional quality adjusted life years and, by doing so, has avoided the issue of valuing time or quality of life in financial terms. Such an approach assumes that a year of life is an additively separable unit of outcome. It has the same value regardless of when it is received or who receives it and is the same regardless of how much total life a person may be expected to live. It does not take into account the additional benefits that might arise from health care, for example, of younger people knowing that they will be cared for later in life, or that less healthy people in society have access to the services they require now. In the context of services that are provided to people towards the end of their lives, the value of these other kinds of benefits may be considerable.

### **Evaluation of complex interventions**

There are particular challenges in undertaking economic evaluation in the context of services that could be described as inherently complex. These are services where the conceptualisation of the welfare benefits derived from the intervention is difficult to define or have various conflicting definitions, and therefore hard to value and measure in the context of research. It also applies to services where the relationship between service inputs and outcomes is not easy to describe, or that incorporate a range of different inputs so that individuals may access a different range or intensity of inputs within the same service. The separate components of the service may not themselves have been well evaluated, and overall, there may be a general lack of understanding (or disagreement) about how services should be provided to maximise their value to patients.

Palliative care is an example of a complex area of health and social care. But the issues raised in the economic evaluation of this service are not exclusive to this type of health care. The

problem of evaluation in this area illustrates issues that are common across complex services. The nature of this complexity is explored in this thesis. Palliative care is designed to be responsive to the needs of individuals and services are tailored to their individual circumstances. The same input may provide a different magnitude and type of benefit depending on who is receiving it. The benefits may be diverse and not easily defined, even by those who provide the services. However, in some palliative care contexts, the benefits are clearer than in others. Acute pain and physical symptoms at the end of life can be reduced with drugs and clinical management, and clearly defined positive outcomes can be measured.

Palliative day care (PDC) is an intervention that aims to meet a wide range of individuals' needs. These individuals will already have their pain and physical symptoms under control (at least to the extent that they are not housebound or receiving inpatient care) and their needs are more difficult to identify since they may relate to a range of different problems to do with how they are coping with their illness and their life circumstances. The purpose of palliative day care is to work with palliative home care services to support patients to live independently for as long as possible with a range of services such as physical therapies, clinical advice, counselling and social support available at a day centre. Some people only use the service infrequently whereas others may need more intensive support. There is also some difference of opinion as to how services should be organised and what they are for.

Clearly, these problems also exist in other services. These are areas of health and social care where there are considerable difficulties in undertaking clinical trials to establish the efficacy of the service and where it would be difficult to conceive of how such a trial could be designed. These are areas where, typically, little evaluative work has been undertaken beyond descriptive analysis. The challenge is to design evaluative studies that go beyond descriptive studies and provide evidence that can be used in decision-making in contexts where evaluative evidence to date has been minimal.

The providers of palliative day care services clearly believe that they provide a valuable service otherwise it could be argued they would not be offering it. Many patients who are referred to palliative day care decide to attend, and attend more than once, so they must do so for a reason. But the value in terms of the benefits gained from attending a centre has not yet been established through evaluation. This means that its value cannot be compared with that

provided by other palliative care services or with the value of using the resources employed in palliative day care in other ways.

It may be the case that the dominant approaches to evaluating health care based on the measurement of life years saved and improvements in health-related quality of life are not sensitive to the nature of the benefits that may be produced by attending palliative day care. The kinds of benefits that may be important in palliative day care might be the social aspect of palliative care, bringing about a change in outlook in people who attend, or making them feel more able to cope with living with advanced disease. While these benefits or changes in welfare may be hard to define (even by the people who provide these services) this does not mean they cannot be very significant changes in welfare to the people who experience them. Furthermore, the value of palliative day care to those who attend may be different from the objectively measurable changes in health gain achieved. The subjective value of a health care intervention could be defined as the value to the individual receiving the intervention, which may differ between individuals. For services such as palliative day care, the value to an individual may be different from the objectively measurable or observable change in health status as perceived by those assessing the individual, for example a health professional. The notion of the 'value' of a health care intervention may therefore differ depending on whose perspective is adopted, that of the individual (defined here as 'subjective') or defined as measurable change in health status (defined here as 'objective'). Instead of considering the value of health care in terms of health gain, defined in terms of additional units of time adjusted for quality, it might be more meaningful to consider what an individual might be willing to give up (in terms of financial sacrifice or consumption forgone) in order to access palliative day care. This argues for the return to the fundamental idea in welfare economics of value measured by the satisfaction of preferences.

In the context of palliative care however the subjective value of welfare and objective measurement of health gain may be at odds. Those attending palliative day care may have strong preferences for attending a centre but objective health gain may not be measurable. In the case of palliative day care, there may be strong preferences in society for the presence of services for people at the end of life with needs that palliative day care can provide. The value of subtle improvements in quality of life when faced with a life-limiting illness or towards the end of life may be greater than the value of measurable changes in quality of life at other times of life.

Furthermore, the provision of these services may be seen as important to society as a whole. This is supported by the fact that hospices within the voluntary sector have been successful at fundraising from the public, even though it is clear that palliative care does not lengthen (and can even shorten) life, and may provide only a few days or weeks of benefits. Therefore there may be something of an anomaly between the value of palliative care services as measured using the QALY or health gain approach, and the support for these services from the general public. This anomaly may not be specific to palliative care, but may also exist in other service. These issues will be addressed in this thesis.

### **Approaches to the evaluation of complex services presented in this thesis**

It is clearly important to develop methods of evaluation that can compare benefits across different types of interventions, and it is necessary to keep the methods as simple as possible for any given health care evaluation problem. Disease-specific outcome measures have been adopted in the evaluation of palliative day care, but not as part of an economic evaluation. They did not show evidence of changes in quality of life, but one argument is that they were not designed for palliative day care settings. They focussed on domains of health related to inpatient care (for example, on acute physical symptoms). Testing the appropriateness of using a global quality of life instrument in the palliative day care setting is one aim of the empirical research presented in this thesis as this has not been done before and it could be argued that a global measure might be more appropriate for an intervention that aims to improve overall quality of life. However there is good *a priori* reasoning that the global quality of life approach will not be sensitive to any changes in quality of life that would be expected from a complex health service such as palliative day care as the levels within each domain are too broad and not focussed on the domains that are most important in palliative day care. Furthermore, it is argued that an approach is needed that reflects the value of these services to the people who use them.

The welfarist framework has made something of a comeback in health economics, despite the practical difficulties of this approach due to perceived difficulties of using the health gain/ QALY approach in some contexts. Other methods of economic evaluation that focus on the satisfaction of preferences as the measure of outcome or 'good' for society may be insightful and are considered in this thesis. Since one of the aspects of complexity in health and social care is its multidimensional nature, it is argued that an approach that can incorporate patients' preferences for different aspects of services rather than the service as a whole could be a useful

approach to adopt. Seeing a complex service as bundles of specific attributes of care may provide insights into which aspects of care are most valued, and which aspects people might be prepared to forego.

One approach that is seen as promising in the context of economic evaluation of health care is discrete choice experimentation, a method in the family of conjoint analysis methods. This type of experiment can estimate individuals' preferences for a range of aspects of care in terms of what other aspects of care (or financial sacrifice) individuals would be willing to trade-off to have these attributes. It has been seen as a useful approach in contexts where there is no one primary health outcome, and as a way of incorporating individuals' strength of preference for different aspects of care into the evaluation process. The constrained choice presented to respondents means that not all aspects of care can be seen as highly valued simultaneously, and people have to make choices about the aspects of care that matter most to them.

In the context of evaluating palliative day care, understanding patients' preferences for particular aspects of care in order to provide services that reflect patients' values may be at least as important as considering the health gain from these services. If patients' strength of preference could be quantified, this might provide evidence on which combination of services would optimise patients' subjective welfare, or which combinations of services are most preferred by different groups of patients (older/younger, socially isolated/with carer support). The satisfaction of the preferences of individuals who have advanced disease and are facing the end of life may be seen as important to the general public as well. Where other methods may not provide any insights at all into how resources ought to be used, this approach may provide the only kind of economic data to inform decision-making.

## **SECTION 2. THE AIMS AND OBJECTIVES OF THE THESIS**

The themes that are developed in this thesis have arisen out of conceptual and empirical work undertaken by the author in the area of palliative day care. Before this work had been undertaken, there had been no published economic evaluations of palliative day care, and little commentary on the challenge for economic evaluation in this area. A previous study undertaken by the author before the thesis had begun highlighted the problems that needed to be addressed. It faced the problem that a full economic evaluation could not be undertaken without a clear indication of the benefits (or otherwise) of palliative day care and that this would be difficult to obtain. Building on this study, the idea for this thesis was to undertake another

empirical study in palliative day care to explore different ways of measuring or valuing the benefits of the service in ways that could be incorporated into an economic evaluation. Two distinct and contrasting approaches were adopted: the first was to use a straightforward health gain method using a global quality of life measure (disease specific measures had been adopted in the previous study) to assess whether it would be possible to derive a quality-adjusted life year to measure the value of palliative day care. Second, a preference-based approach was adopted, by undertaking a discrete choice experiment in the same study. The aim of the discrete choice experiment was to try to measure the relative value of different aspects of palliative care to those who attended a day care centre. It was not clear at the beginning whether a willingness-to-pay component could be incorporated into the experiment, and this question explored in the thesis.

The aim of the thesis therefore is to contribute to the understanding of how to undertake economic evaluation in areas of health and social care that are inherently complex and difficult to investigate. It traces the roots of welfare economic thinking back to the foundational ideas of utilitarianism and explores the purpose and application of economic evaluation arising from welfare economic principles. The challenge of applying welfare economic approaches in health and social care contexts is examined and current debates about how to conceptualise and measure benefits are presented. The argument is that the measurement of outcomes is difficult generally, and that the measurement of outcomes of services that are complex presents even more difficulties. The empirical evidence that is presented is an attempt to make some progress in this area and to find a way of presenting economic data that is useful for decision-making.

The definitions of palliative care and aims of palliative day care are presented in the next section to provide a contextual backdrop to the rest of the thesis. It also situates the evaluation of palliative day care in the context of current UK government health policy initiatives to bring statutory and voluntary sector services more closely together. It considers the tensions that this brought about by the necessity to demonstrate 'value for money' compared with other services. These tensions are pertinent to the thesis since they highlight the need to find ways of presenting the benefits of palliative day care in ways that are fair and at the same time not over or under-emphasising the value of these services to the people who use them.

### SECTION 3. INTRODUCTION TO PALLIATIVE CARE AND PALLIATIVE DAY CARE

Palliative care is “the active total care of patients whose disease is not responsive to curative treatment. Control of pain and of other symptoms, and of psychological, social and spiritual problems is paramount” (WHO 1990). Palliative care is also seen as unique in considering quality of death alongside quality of life (Jarvis et al. 1996; Whynes 1997, Wallston et al 1988). The focus of palliative care is defined primarily on comfort and support, and is a comprehensive, interdisciplinary service for the care of patients and families facing a life-limiting illness (Billings 1998).

Some aspects of palliative care are also applicable to patients in earlier stages of illness. The National UK Council for Hospice and Specialist Palliative Care Services has set out the stated aims of palliative care services in the UK, outlined in box 1.1.

In the UK, there were between 100,000 and 120,000 patients with cancer who have palliative care needs in the 1990s (Department of Health 1995). This figure does not include patients with other diagnosis who have similar needs. The Expert Advisory Group on Cancer defined palliative care not only as terminal care, but patients may have palliative care needs at the time of diagnosis and should therefore be an integral part of cancer care. (Department of Health *ibid.*)

#### **Box 1.1 Aims of palliative care (National Council for Hospice and Specialist Palliative Care Services, 1997)**

**Palliative care:**

Affirms life and regards dying as a normal process

Neither hastens nor prolongs living

Provides relief from pain and other physical distress

Integrates psychosocial and spiritual aspects of patient care

Offers a support system to help patients live active lives as long as possible

Offers a support system to help families cope during patients' illness and in their own bereavement

Many individuals who are referred to palliative care or hospice services have had a cancer diagnosis (Higginson 2000). However, patients also attend who have other forms of illness such as HIV/AIDS, motor neurone disease (Higginson *ibid.*). It has been reported that there is now a “significant investment” in palliative care and a high level of spending on palliative care which is

undertaken by people who are not palliative care specialists, estimated to be about £500 million per annum (Tebbit 1998).

There is a variety of services available for people with palliative care needs in the UK, with key resources identified as hospices, specialist nurses working in community and hospital teams, Marie Curie nurses who provide night sitting and other support services, and oncology and palliative care inpatient and outpatient services (Appleby 1994). These services are offered alongside general clinical services such as primary care (general practice, district nurses, and community pharmacies) and other statutory community services. Nursing homes also form part of the package of palliative care since patients may stay in a nursing home while recovering from a hospital stay

In 1998, in a report by The National Association for Health Authorities and Trusts, the first recommendations were made to health authorities that palliative care services should be provided in all localities (NAHAT 1998). The recommendation was that services should begin by providing specialist palliative care nursing, then palliative day care, and only then to provide inpatient facilities if local needs could not be met by these first levels of palliative care. However, palliative day care has been slower to develop than inpatient services (Higginson, Hearn et al 2000).

Hospice care has usually been offered outside mainstream statutory health care. As such, they are often described as providing care that is qualitatively different from NHS and social services (Torrens 1984, Abel 1986, Neale and Clarke 1992, Payne 1996). The development of palliative care within the voluntary sector has occurred at the local level in a fairly autonomous environment (Clarke, Neale et al 1995), and with relatively little control by central government or the NHS (Department of Health 2000). The Calman-Hine report indicated that hospices had developed in a random fashion in the UK, reflecting the development of other services in the voluntary sector, and in the early days of the movement, services developed with little consultation or contact with the NHS or other statutory services (Robbins 1998). It has been argued that the input (or control) by voluntary organisations has led to some diversity in the developments, and distinct philosophies of provision are visible (Spencer and Daniels 1998). Another reason for the diversity is the independence of the funding from the NHS. Funding for the different services may come from different health and social care budgets, and different rules apply regarding patients' access to services. Palliative care is therefore provided in a



complex “mixed economy” (Tierney and Sladden 1994). The reliance of palliative care on at least some voluntary fundraising has contributed to services being perceived as “different” from other health and social care services offered by statutory bodies (Department of Health 1996). The range of services also offered in the hospice setting is also dependent on the availability of local volunteers to offer specialist therapies, supportive services, and on general help.

In the past, palliative care services were run autonomously by voluntary organisations and have therefore avoided some of the financial and service development pressures that other sectors have faced (Department of Health 1998). However, primary care trusts can now purchase hospice care from the voluntary sector and there are now fully funded NHS hospice services in areas where voluntary-sector hospices have not been built (Higginson, Hearn et al. 2000). As hospice services have grown, and the volume of funds from government sources has increased, it has become clear that the Department of Health is taking an increasing interest in their funding and quality.

The publication of ‘A Policy Framework for Commissioning Cancer Services: Palliative Care Services’ (Department of Health 1996) spelt out the domains of palliative care that have led the way in introducing concepts of health care which are now being seen as important for a whole range of conditions: the focus of quality of life; the whole-person approach; care which encompasses family and friends; respect for autonomy and choice; an emphasis on open and sensitive communication.

Following this, the National Cancer Plan recommended that the NHS and the voluntary sector ought to be working more closely together and funds have been pledged to hospices to support their activities:

*“By 2004, the NHS will invest an extra £50 million in hospices and specialist palliative care. The Department of Health will agree with the voluntary sector the core services that should be available, so that more patients will have access to these services, and the NHS will make a more realistic contribution to the costs of voluntary hospices. NHS and voluntary sector services will work more closely together.”* (Department of Health 2000)

The Health Act (Department of Health 1999) created new working relationships to enable (the old) health authorities and councils to improve services at the interface of health and social

care. Local Strategic Partnerships were established across the country in 2001 as umbrella partnerships with the remit to improve quality of life and governance in a particular locality through bringing together public, private, voluntary and community sectors within which new National Service Frameworks (NSF). The National Service Framework for Older People is only the third NSF to be developed which indicates the perceived political importance of improving services for this group (Department of Health 2001a). Palliative care services are used mainly by older people so the consideration of quality of care at the end of life is given a high profile in the NSF. The focus on end-of-life care in this document expressed the key values promoted by the palliative care movement, that is, pain and symptom control, social care, spiritual care, complementary therapies, psychological care, communication and bereavement support. This indicated the extent to which palliative care philosophies have permeated into the mainstream policy agenda (if not into practice).

The consultation paper on the NSF for older people was published in January 2002 and it focused on how supportive and palliative care should be defined and the principles that should be adopted. It highlights the difficulties of conceptualising an overarching definition of care for a service for people who have different and changing needs over time. The National Institute for Clinical Excellence has now been commissioned by the Department of Health to provide evidence based guidance on supportive and palliative care and will report its findings during 2002/3 (Department of Health 2001b).

These developments in defining and agreeing how palliative care services should be delivered imply that they will be judged in similar ways to other services. The Department of Health will want to be shown evidence of the effectiveness of palliative care services. Back in the 1980s, the argument was set out by the hospice movement that palliative care was better than traditional or conventional care for people with palliative care needs but there had been little systematic evidence for this up until then (Seale 1989, Torrens 1985, Lancet Editorial 1986). Evaluative studies that consider the outcomes of hospice care are emerging but the evidence is not conclusive and the methods of evaluation have not been very robust. Systematic reviews of studies of palliative care studies have not provided unequivocal evidence of the superiority of palliative care over conventional care (Goodwin, Higginson et al 2002a). There is little published evidence of the cost-effectiveness of palliative care (Bruera and Suarez-Almazor 1998).

It is clear that there is a need for palliative care to present its contributions in ways that allow comparison with other social and health services that compete for scarce resources (Webber 1996, Normand 1996). For aspects of palliative care that focus on pain control and symptom management this is relatively uncomplicated as similar outcome measures as those used for cancer therapies might be adopted (Brown and Sculpher 1999). These are areas where clinical guidelines and technology assessment face the same kinds of well-documented evaluation problems as other types of health care (Birch and Gafni 2002). It is less straightforward for interventions that meet more ill-defined, contested, or non-clinical aspects of care (Bosanquet 1997). This challenge is addressed in this thesis.

### **Palliative day care**

Palliative day care is considered in this thesis as an example of a 'complex' intervention. The nature of this complexity is explored and defined further on. It is a specialist service for people who usually have advanced life-limiting illness and are usually able to live at home. They are perceived to have particular needs that can be met by attending a specialist day centre (Fisher and McDaid 1996, Tebbit 1999, Higginson et al 2000) and crosses the somewhat artificial boundary in the UK between health and social care (Robbins 1998).

Palliative day centres are usually attached to inpatient hospices and patients may move between these settings as the severity of their illness changes over time. Most centres have a main social room, kitchen facilities, and therapy/ consultation rooms. A centre might also have a specially adapted bathroom/jacuzzi, arts facilities, hairdressing facilities, quiet rooms, office space, and access to other parts of the inpatient unit, such as the smoking room and garden. The space emphasises a non-clinical environment and is usually decorated with arts and crafts made by people who attend, as well as work by local artists. The atmosphere is informal, and people are encouraged to socialise and take part in group activities if they want to. Activities are either undertaken in the main room as a group (such as creative activities, relaxation classes, or social events) or are on an individual basis in the therapy/ consultation rooms, bathroom or hairdressing suite. There may also be talks or entertainment by outside visitors or occasional outings, similar to other social care centres.

People who attend usually choose to have one or two therapies or consultations during the day, and the rest of the time is spent in the main room socialising with others, taking part in group activities or resting. Lunch is also served in the main room, and in every centre everyone is

offered drinks from a mobile bar before lunch, which also emphasises the non-clinical philosophy of care. The centres are run by a full-time day care leader with help from other assistants, health professionals, therapists and others (such as religious counsellors) who may work in the inpatient, day care and community settings. Volunteers act as general helpers and may provide transport to and from the centre, depending on the numbers available. Like inpatient hospice care, the input of voluntary support has led to some diversity in the developments, and distinct philosophies of provision are visible (Douglas et al 2000, Spencer and Daniels 1998).

The objectives of palliative day care have been grouped into five themes outlined in box 1.2 below:

**Box 1.2 Objectives of palliative day care (Neale and Clark 1992, Scottish Partnership Agency 1995),**

1. Physical well-being:  
optimal control of physical symptoms and maximal functioning;
2. Psychological well-being:  
improving strategies for dealing with stress, identifying depression and anxiety and increasing self-esteem and confidence;
3. Social support:  
relief of social isolation and provision of respite for carers;
4. Communication and co-ordination:  
ensuring optimal communication between patients, carers and professionals, and between different services;
5. Existential health: engendering hope and helping people to finding meaning and purpose in their lives.

Patients who access palliative care services are identified as having a spectrum of needs or one specific need that it is proposed can be met or helped by a combination of health, social, psychological and spiritual activities provided in the palliative day care centre (Higginson 1993). In common with inpatient hospices and community palliative care teams, palliative day care integrates physical care with support for emotional, social and spiritual well-being (Hearn and Myers 2001).

There are differences in the way that services are provided across different centres (Copp, Richardson et al 1998). Some centres encourage patients to stay all day, others set up appointments for patients who might stay for a while before and after their appointment but then

go home. There is some debate about how much medical support is required in a PDC centre (Edwards et al 1997, Tookman and Scharpen von Heussen 2001), with some centres offering appointments and medical review, with others only providing medical back-up in emergencies. In the United Kingdom in 1991 there were 151 palliative day care centres attended by approximately 4500 patients each week (Eve and Smith 1994). In 1996 there were known to be 230 day care centres accepting patients for one or two days a week. By August 2001 there were 260 different palliative day centres in the UK and 39 in the South-East region (personal communication, Hospice Information Service 2002). Provision of palliative day care has expanded rapidly, ensuring both that there is a service available in most parts of the UK, alongside extensive developments of day care and home care teams.

This presents a challenge to evaluators, both in terms of determining the boundaries of the intervention that is being assessed, and in terms of defining the outcomes of the intervention. This has to be done in a way that is seen as appropriate to those attend PDC, meaningful to policy-makers, and valid and reliable in the eyes of the research community.

The provision and evaluation of palliative day care will be explored in more detail in chapter 3 in light of the evaluative studies and qualitative research that has been undertaken to inform the economic studies presented as part of this thesis.

#### **SECTION 4. THE STRUCTURE OF THE THESIS**

The next chapter introduces the theoretical foundations on which the thesis and empirical work is structured. It provides a discussion of the theory of welfare economics and considered different methods of economic evaluation and why they have been adopted in different contexts. The purpose of this is to understand the roots of economic evaluation, the theory on which it is based, and why economic evaluation has developed the way that it has. This will provide a foundation for understanding why there may be a particular problem in the measurement of complex health and social services. It is argued that some of the problems of not adhering to welfare economic theory are particularly problematic in areas where the outcomes of interventions cannot be easily measured in terms of health gain, or process measures. The approaches to economic evaluation adopted in health economics are also reviewed. The discussion examines the extent to which these methods have developed within or moved away from the central assumptions of welfare economic theory. By moving away from preference satisfaction to health gain, these approaches are pragmatic and have worked well in many

contexts, but when it comes to the evaluation of more complex forms of care, they may not be so helpful.

Chapter 3 considers the current state of economic research in palliative care as an example of a complex area of health care. The evidence is assessed in light of the debates set out in chapter 2, especially with respect to how the outcomes of palliative care have been defined and measured. Important gaps in the literature are identified in this review of the economic evidence. In light of this, the practical and ethical challenges of undertaking studies in complex settings are considered.

There is a body of evidence in the qualitative literature that provides insights into the nature of palliative day care, and this evidence is important for the design of the choice experiment. Evidence from a review of studies of day hospital and adult day care and of social services for older people is also presented as these studies may provide insight for the evaluation of palliative day care.

Chapter 4 builds on the evidence presented in chapter 3 and considers the arguments for undertaking simple quality of life evaluation alongside the arguments for the adoption of preference-based evaluation methodologies. Both approaches are undertaken in the empirical investigation of outcomes for this thesis. The EQ-5D quality of life instrument is introduced, and the methods explored. The choice experiment approach to valuing outcomes is also presented in this chapter, along with a detailed discussion of the methods of undertaking choice experiments, and the theoretical underpinnings of this methodology, since it is less well known than the EQ-5D approach. The extent to which this methodology addresses some of the problems inherent in the health gain approach is explored, and the possibility of deriving monetary valuations of outcomes is considered.

Chapter 5 provides a detailed account of the methods used in the empirical study and the design of the choice modelling experiment that was undertaken. The EQ-5D study was undertaken as part of a larger evaluation of PDC that included an evaluation of the costs of palliative day care, but the detailed methods and results are not reported since they are not included within the overall aims of the thesis (a summary is attached at Appendix A). The choice modelling experiment was designed in light of the qualitative descriptive evidence of palliative day care and previous experience of trying to undertake an economic evaluation in this

context. This chapter describes how this study was designed from the description of the nature of the research problem, to the final lay-out of the experiment. It refers to other studies that have used these methods in economic evaluation and explores the extent to which this study emulates these studies and how it differs from them.

Chapter 6 presents the results of the two empirical studies, starting with the EQ-5D study and then presenting the results of the choice experiment. The results of the choice experiment are reported following the analysis plan outlined in the previous chapter.

Chapter 7 discusses the important findings of the empirical studies presented in chapter 6 and considers what this knowledge contributes. The discussion in this chapter focuses on the contribution of this empirical work to decision-makers in areas where other forms of evaluation have not been insightful. It appraises the strengths and weaknesses of the empirical studies and considers whether the quality of life approach has provided sufficient evidence to support this approach in other contexts where outcomes are not clearly defined. It then assesses whether the choice experiment approach has provided any new or alternative insights into the value of complex services.

Chapter 8 is the conclusion of the thesis. It revisits the main themes set out in the introduction, and reviews the progress that has been made in tackling some of the problems set out at the beginning. It assesses what the overall findings mean for policy, and the methodological contribution to the evaluation of complex services. It then presents the overall contribution of the thesis and considers the shortcomings both of the empirical study and of the thesis as a whole. Finally, it signals areas for future research leading on from this thesis in palliative care research, and in the development and validation of theoretically sound economic evaluation methods that measure complex outcomes.

## Chapter 2

## Theory and application of welfare economics and economic evaluation

### Introduction

This chapter provides an overview of the theory of welfare economics as the cornerstone of economic evaluation of health care. The focus of the discussion is on understanding why economic evaluation methods have developed as they have and how they relate to the underlying theory. The purpose is to consider the appropriateness of different methods for evaluating complex services, and why some methods might be more relevant than others.

The theoretical overview in the first section is a discussion of the intellectual origins of welfare economics and the nature of the problems that welfare economics has addressed. The origins of this work can be traced to the ideas of 19<sup>th</sup> century economists and even further back to Adam Smith's 'The Wealth of Nations'. The discussion highlights the key theoretical assumptions of welfare economics and the ways in which these assumptions have led to both theoretical and intellectual developments in health economic evaluation in the 20<sup>th</sup> and 21<sup>st</sup> centuries.

The second section of this chapter considers the methodological developments in the field of economic evaluation in health economics in particular, and the varying importance of welfare economics in these developments. It reviews the debates about how to value health care, and how far society's preferences (as opposed to objective measures of health gain) have been included in methods of evaluation. The different methods are considered in the light of their application to complex services to assess which would be appropriate in these contexts.

### SECTION 1. THE ORIGINS OF ECONOMIC EVALUATION IN THE WELFARE ECONOMICS TRADITION

The forefathers of modern welfare economics were moral philosophers as well as economists. The two disciplines were intrinsically related to one another, involving political and moral thinking, relating values and how society ought to be (Edwards 2001). Adam Smith's 18<sup>th</sup> century theory of *laissez-faire*, where egotistical individual action leads to social good, is a founding theory. The axioms of this theory are interconnecting neo-classical ideas of constrained choice, consumer sovereignty, consistency of tastes, perfect knowledge and



perfect mobility of factors. This has not changed in neo-classical economics since that time. (Blaug 1985, Edwards 2001).

### Utilitarianism

The theory of *laissez-faire* or market economics was seen as an alternative to theories of natural and divine law for defining the right way to live (Cole, Cameron and Edwards 1983, Salter 1994). Jeremy Bentham, a philosopher with an interest in economics, first proposed that consumer satisfaction was measurable. He proposed that all human beings have roughly equal capacity for pleasure and pain; therefore equality in consumption was desirable for the greatest happiness of the greatest number (utilitarianism).

But the psychological foundations of utility theory were crude. Bentham's work relied on the theory of 'psychological hedonism', that all behaviour was motivated by the pleasure/pain principle. This theory provided a rationale for consumer behaviour but the underlying hedonistic assumptions were strongly challenged in the face of the psychological evidence for how people actually behaved. This was summarised succinctly in the 1920s by the psychologist William McDouglas who wrote:

*"It would be a libel, not altogether devoid of truth, to say that classical political economy was a tissue of false conclusions drawn from false psychological assumptions."* (McDouglas 1923, quoted in Landreth and Colander 1994)

In a summary of the work of Bentham, one author has suggested that Bentham sometimes referred to utility as if it were a state of mind related to pleasure and pain, at other times as if it were a metaphysical property of the objects that produce pleasure or pain (Bonner 1995). Bentham's theory of value is a more complex analysis than is commonly known in that he acknowledges that intensity, duration, certainty, and nearness/remoteness factors are also included in valuing pleasure and pain. Bentham also recognised that losses are more keenly felt than future gains and that individuals differ in their capabilities. He proposed that all policy should be assessed on the basis of its effects on human feelings, and that there were trade-offs between efficiency, equality and happiness (Bonner *ibid.*).

Bentham was highly influential in promoting the ideas of utilitarianism. Bentham believed changes in utility would eventually be measurable so that economics would have the same precision and objectivity as the natural sciences (Mirowski 1989, in Ackerman 1997). Social utility or social welfare was simply the sum of individuals' utility. However, although utility was defined as a quantitative magnitude, it could not be either directly observed or

measured. It was conceived as a latent property that could only be measured by observing its impact on (utility-maximising) human behaviour. This is a fundamental unchanged property of utility as defined in neo-classical economic theory today (Robinson 1964).

Utility was therefore conceptualised as an accounting unit by which different combinations of goods and services could be made commensurable and evaluated by the individual. Pleasure could be increased (or pain reduced) by trading and consuming different bundles of goods and services that contained different levels of utility, as perceived by the individual doing the trading. However, while proposing that utility could be measured this way, Bentham never proposed a practical method of empirical measurement of utility. Smith's *laissez-faire* economics had demonstrated to his satisfaction that utility is maximised by free competition and trade, and this neatly avoided the problem of having to measure it.

There was apparently little interest in the theory of utility until John Stuart Mill proposed a more complex approach to understanding human motivation than individual rational hedonism (Hunt 1979). Mill rejected two of Bentham's central tenets: that all motivation could be reduced to self-interest, and that the individual is always the best judge of his own welfare. He also challenged the idea that different forms of pleasure could be summed across individuals or within individuals. His idea was that different sources of pleasure differed in quality and were not therefore commensurable units of utility. Some forms of utility were more worthy, or more valuable to society, than others. He rejected the extreme values of *laissez-faire* by observing the failures of capitalism in his day. He proposed that where markets failed, governments should intervene in order to promote more equitable distribution of income and to assist the poor. This difference in moral philosophy from Bentham is a fundamental "intellectual schism" between two pathways of economic thought within neo-classical economics. The idea that utility maximisation should incorporate some additional moral dimensions, such as equity considerations, rather than being simply the sum of individual utility is an important theoretical concern in modern applications of economic evaluation. This is especially true of economic evaluation of health care.

The next generation of economists continued this intellectual schism in the theory of utility (Black 1990). Jevons and Marshall had different understandings of utilitarianism, representing different approaches to welfare economics. Jevons' work was based on Bentham's theories and he opposed the re-working of utility by Mills that some forms of utility were more or less worthy than others. He proposed an alternative "ordinalist" view of utility that was the foundation of the ordinalist revolution achieved by the work of Pareto (Ackerman 1997). Jevon's work relied on Bentham's one-dimensional net sum pleasure/

pain principles. But he proposed that measurement of utility was nearly impossible and that utility and disutility were hypothetical constructs. This approach rejected the idea that utility could be measured or compared between individuals. It emphasised the unpredictable diversity of individual desires rather than their commonality. The work of Jevons (followed by Fisher and Pareto) demonstrated that the technical theory of human behaviour could be developed without cardinal measurement or interpersonal comparisons of utility.

Marshall on the other hand made a distinction between ethical and economic arguments (Ackerman 1997). He proposed that utility was measurable in terms of money, but rejected the “crude” hedonic demand principles as the theory of value. He suggested that human motivation was more complex than a simple ledger of pleasure and pain. Marshall was influenced by Mills in his philosophy and economic theory. However he acknowledged that non-material values were not amenable to empirical analysis and therefore were not appropriate subjects for economic enquiry. These ideas became detached from the technical aspects of work by the same authors (Ackerman *ibid.*).

Jevons and Marshall proposed different future directions for economics: Jevons favoured economic sub-disciplines and the development of a science of economics, akin to the natural sciences. Marshall favoured a synthesis of disciplines and argued for a broader understanding of human behaviour but in quantifiable terms, set against the measuring rod of money.

### **Vilfredo Pareto's contribution to modern welfare economics**

The fundamental principles of economic evaluation in health care are grounded in Paretian welfare economics since Pareto was the economist who set out the strict principles on which comparisons between alternative policies could be made in the utility maximising framework. Adam Smith had already addressed the question of human welfare by proposing that welfare could be maximised by the invisible hand of markets and trade. However, the assumptions that underpin the economic theory (and its limitations) were made explicit by Pareto.

Pareto set out principles of welfare economics in the late 19<sup>th</sup> and early 20<sup>th</sup> century. He was working in a general equilibrium framework developed by Walras, an economist concerned with the problem of whether it was possible to define a finite level of production and consumption where the economy would be optimised. Pareto considered the formal conditions for optimal levels of production and consumption, which he argued would lead to the best conditions for the welfare of individuals living in that society. The analysis was

based on neo-classical economics, but its broader concern with the social benefits of resource optimisation led to a new branch of economics known as welfare economics (Jackson 1992, Landreth and Colander 1994). It is interesting to note that Walras (who taught Pareto) was the founder of axiomatic mathematical analysis of competitive equilibrium, and he drew sharp distinction between applied (market) economics and what he called “social” economics.

Pareto was a follower of Walras’ mathematical theory. Pareto’s contribution was to give a formal definition to the concept of social welfare and the conditions necessary to achieving a socially optimal level of welfare. This assumed (as for neo-classical economics in general) that individuals are rational, welfare-maximising decision-makers, and who make decisions to trade their wealth/income for goods and services that improve their well-being (Winch 1971).

Pareto started to formalise the conditions, known as the ‘marginal conditions’, of welfare maximisation, for achieving optimal resource allocation of society’s resources (Brown and Jackson 1978). Under these conditions, a Pareto Optimum is one where the well-being of all individuals is maximised so that any economic change would make at least one person worse off. A Pareto Improvement in an economy, therefore, is one where any reallocation of resources makes at least one person better off without making anyone worse off. A move towards these optimising conditions could be said to be an improvement in social welfare. The marginal conditions are also the conditions necessary for perfect competition in classic economic demand theory. Therefore, these conditions pertain to a theoretical model rather than the conditions that are comprehensively attainable in real life (Dasgupta and Pearce 1972).

An important break from past debates on social welfare (the “Old” welfare economics) was Pareto’s rejection of the idea of interpersonal comparisons of utility and cardinal measurement. Until Pareto, it had been accepted that individual utility could be summed to calculate total social welfare (Blaug 1985). However, Pareto argued that the utility (or the strength of preference) derived from a commodity could not be assumed to be the same for two individuals and therefore could not be summed to produce a total amount of utility. This was known as the ‘ordinalist revolution’. The only type of approach that would not rely on making comparisons between individuals’ welfare was a straightforward ordinal ranking of preferences. This meant that it was possible to know the ranking of preferences between individuals, but not the relative strength of preference between them. Within this framework, it could be determined that a person prefers A to B but not how much more A is preferred to

B. If some people prefer A and some people prefer B, it is not possible to say which option has the strongest support overall.

Pareto was concerned to maintain economics as a positive rather than a normative science and as a consequence he narrowed the criteria for what could be known about individuals' and society's welfare. This meant that there were only very limited circumstances in which it could be said categorically (or positively) that a welfare improvement had been achieved. To return to the example above, it would only be true to say A represents an improvement in social welfare if one person or more preferred A to B and everyone else is indifferent between A and B. In other words, the only changes that could be considered an improvement in social welfare are those where at least one person is made better off and no one made worse off.

It has been argued that Pareto's work is grounded in a school of economic thought that considered economic theory in isolation from the practical reality of decision-making (Mishan 1988). His work is more comfortably situated in the mathematical rather than the ethical schools of economics. But 20<sup>th</sup> century economists have argued that the conditions for Pareto optimality were anything but politically neutral because of the assumptions about human behaviour on which his ideas were based (Sen 1970a).

Pareto's analysis might have fallen into obscurity due to the fact that so few real life situations exist where a Pareto Improvement might be determined. In nearly all policy changes there will be winners as well as losers. In his context, Pareto does not provide guidance since it is not possible to determine whether the winners win more than the losers lose.

### **Later developments of welfare economics**

Pareto never actually finalised the formal rules of Pareto optimality. This was undertaken in the 20<sup>th</sup> century. Later economists distinguished between value judgments and factual analysis, minimizing the importance of the problems of interpersonal comparisons of utility. The results of this work were formalised in the 1950s by Arrow and Debreu who set out the two fundamental theorems of welfare economics (Arrow 1963). These were conditions that would have to meet for a Pareto optimum to be realised<sup>1</sup>. These conditions made reaching a Pareto Optimum a theoretical concept rather than an achievable goal since the number of situations in which there are only gainers is minutely small (Jackson 1992).

---

<sup>1</sup> The first theorem states that a competitive equilibrium is also a Pareto Optimum. The second theorem states that any Pareto Optimum can be achieved via competition if the appropriate lump sum taxes and transfers are imposed on firms.

Before Arrow and Debreu, a second generation of “new welfare economists” of the 1930s had attempted to make some progress beyond this theoretical *impasse* of the impossibility of interpersonal comparisons of utility that had reduced Pareto’s influence to almost nothing in his own time (Jackson 1992). Pareto’s work was rescued by developments in welfare economics that proposed the concept of gainers compensating losers in policy change.

Pareto himself acknowledged the fact that if the gains from an economic change were greater than the losses, then social welfare had increased, but he never was able to formalise this and maintain the strict adherence to positive economics. One of his contemporaries, writing in 1908, Barone set out more formally the conditions under which gains might outweigh the losses (Landreth and Colander 1994). Then in 1939, two economists, Kaldor and Hicks - separately from each other - considered the idea of a Pareto improvement where gainers compensated losers. The Hicks-Kaldor criterion stated that if the gainers compensated the losers up to the point where losers returned to their original welfare position (i.e. were indifferent between their original welfare position and their new position with compensation) and there were still overall gains to gainers, then social welfare had increased. This was a new foundation on which welfare economics could be based, and came to be known as the ‘second welfare economics’ (Hicks 1975).

The decision-rule was later refined since in many situations the costs of identifying the gainers and losers could outweigh the final benefits and make the change not worthwhile. The Hicks-Kaldor criterion evolved into a *potential* compensation by gainers to losers without any exchange of wealth necessarily taking place. This meant that Pareto’s economic theory could be translated into a practical analytical tool. It still avoided the ethical question of who gains and who loses from any intervention or economic change. But as a tool for assessing overall costs and benefits before the issue of distribution arose, it was judged to be useful. It became the foundation of cost-benefit analysis.

While there were important improvements and refinements to Pareto’s welfare framework in the mid 20<sup>th</sup> century, there were important critiques of the framework. These arguments brought about a decline in the application of the welfare economic theory to real world problems (Ng 1990). Two of these critiques were Arrow’s general possibility theorem and Scitovsky’s double criterion principle. Arrow’s theorem (referred to as the third fundamental theorem of welfare economics) proved that under a few simple assumptions there is no logically consistent, non-dictatorial social welfare function that ranks all social outcomes (Hammond 1987). It proved the impossibility of adding individual ordinal utilities into a social welfare function (MacKay 1980). The second criterion, also called the reversal paradox,

challenged the basis of Pareto's criteria that a change was always for the better if it made one person better off without making anyone else worse off (Scitovsky 1973). What he demonstrated was that if an allocation A is deemed superior to another allocation B by the Kaldor-Hicks compensation criteria, by a subsequent set of moves by the same criteria, it could be proved that B is also superior to A. These critiques threw much of welfare theory into confusion during the 1940s and 1950s and imposed huge problems in the practical application of use of welfare economics.

The overall aim of welfare economics has been the maximisation of the social welfare function derived from individual desires (Arrow 1963). Even with the theoretical challenges set out above, economists have adhered "with tenacity" (Edwards 2001) to the underpinning value judgements of Pareto welfarism. These are that every individual is the best judge of his own utility and each has a specified utility function. Social welfare is defined as the sum of individuals' utility. What has been adhered to with less enthusiasm is the notion that the utility of individuals cannot be compared (Arrow 1963). One of the reasons for this has been the problem of what decision rule could replace Pareto optimality (Edwards *ibid.*). This problem will be returned to in later sections in the context of health economic evaluation.

Late twentieth century economists have argued that welfare economics should be rejected since it does not provide practical, value free decision rules. Sen has shown that Pareto optimality assumes that if a move makes everyone better off, society is automatically better off. This may be an unobjectionable value judgement for many, but it is a value judgement nonetheless (Sen 1987). There is also the problem that using Pareto optimality as a criterion to determine welfare isolates Hume's dictum that you cannot derive a 'should' from a fact. Therefore, in its pure form, Pareto optimality may have little to contribute to real world decision-making. Almost all social policy decisions will inevitably hurt some people, even if in a small indirect way, while helping others.

In the wake of Arrow's general possibility theorem, and other critiques, a new approach to the problems of welfare economics had to emerge. Social choice theory set out to examine the manner in which individual choice, preferences and welfare should enter into social judgements and decisions about economic matters. This theoretical development coincided with philosophical discussions of equity and ethics in economics in order to expand the subject matter and range of welfare economics.

Sen has been very influential in this field. Sen appeals to the common sense belief that people's experience and satisfactions are comparable and human beings act *as if* they could

compare their satisfaction with others. Sen has analysed Arrow's theorem and has questioned the strict axioms he presents, especially the rejection of interpersonal comparisons of utility (Sen 1970a). Rarely are decisions made *without* some interpersonal comparisons of utility. Furthermore, he has argued that satisfaction of private preferences alone is an inadequate basis for social judgement. Therefore at the heart of Pareto's rules is a "liberal paradox." (Sen 1970b). This paradox is that Pareto decision rules are incompatible with even the most minimal interpretation of individual rights. Sen has suggested some modification of the Pareto rule so that an individual's choices for the satisfaction of personal preferences can be separated from an individual's preferences to be counted for social choice.

### **Developments in utility theory - axiomatic utility theory**

Mathematical interest in utility theory as a theory of human behaviour and motivation has developed alongside the theoretical debates about the usefulness of Paretian welfare economics. Its contribution is in developing the underlying theory and demonstrating a relationship between human behaviour and the otherwise obtuse concept of value and utility. This is a valuable contribution as it can get around some of the problems of Paretian welfare economics of measuring utility, described earlier.

Axiomatic utility theory starts from the point of individuals' preferences. The theory proves that, provided preferences conform to some axioms, they can be represented by a utility function. Of two alternatives, the one that is preferred will have the higher utility value. Values taken from this function are called utilities. In any choice content, when faced with a set of alternatives, the utility function assigns a greater utility to the alternative that is preferred. So utility has adopted the meaning "the value of the function that represents a person's preferences" (Broome 1999). The theorem of utility theory is that a person always prefers what is best for them and a person always prefers exactly what is in their own self-interest has been criticised as not taking into account human traits such as altruism and imprudence. The debate is whether this meaning of utility rather than other concepts of value or "need" are useful when applied to non-market commodities, such as health care. This represents an important divergence of methodology in economic evaluation, as will be discussed later in this chapter in considering health economic evaluation.

However, axiomatic utility theory itself does not make assumptions about the self-interest of individuals or why they make the decisions they do, and whether they are altruistic or selfish. It is only concerned that individuals conform to specific mathematical axioms of behaviour.



One development of axiomatic utility theory has become the foundation of economic evaluation theory is expected utility theory (EUT).

### **Expected utility theory**

EUT has its modern roots in the mathematical decision theory developed in the context of game theory in the 1940s by von Neumann and Morgenstern. However, its historical roots go further back. Mathematicians in the 17<sup>th</sup> century assumed that a rational individual would choose to yield the highest possibility utility (in terms of money) when faced with a choice of alternatives (Oliver 2002). Later in the same century, it was argued that the value of wealth increases at a decreasing rate, and that utility is evaluated not in terms of expected value, but in terms of expected subjective value to the individual. This was a cornerstone of expected utility theory.

The work of the founders of game theory provided the first formal exposition of the axioms of EUT as a way of considering utility that overcomes some of the problems of Pareto optimality and welfare economics (von Neumann and Morgenstern 1944). The theory of behaviour towards risk first set out by von Neumann and Morgenstern is essentially the hypothesis that the individual possesses (or acts as if possessing) a utility function that adheres to a set of strict axioms of rationality. This means that, when faced with alternative risky choices, the individual will choose the alternative that maximises expected utility. Since the outcomes being considered could be alternative levels of wealth, years of life, commodity bundles, or even non-numerical consequences (a healthy baby, for example), the approach is open to a wide range of applications and choice situations. Much of the theoretical work in the economics of uncertainty is undertaken within the EUT framework.

EUT assumes that individuals have a preference ordering over objects that may be represented by a preference function. A particular alternative in a choice presented to an individual will only be preferred by that individual if it is assigned a higher value in the individual's preference function (and vice versa). There exist dozens of formal axiomatisations of EUT in its different contexts. Most start with four basic axioms: completeness, transitivity, continuity and independence (Box 2.1).

The theory demonstrates that, if the axioms are adhered to, there exists a utility function, where, like utility theory under certainty, the alternative chosen by an individual will have higher utility, but in addition, it shows that the utility of an alternative is also its *expected* utility based on probabilistic outcomes. This means that an individual who adheres to the

axioms of EUT will choose the alternative that has the best probabilistic outcomes under uncertainty.

**Box 2.1 The four basic axioms of Expected Utility Theory, (as described in Oliver 2002)**

If preferences are **complete**, then either  $Y_1 \geq Y_2$  or  $Y_2 \geq Y_1$ , where  $\geq$  is the preference relation "at least as good as"

If preferences are **transitive**, then for  $Y_1 \geq Y_2$  and  $Y_2 \geq Y_3$ , then  $Y_1 \geq Y_3$

**Continuity** requires that if  $Y_1 \geq Y_2 \geq Y_3$ , then for some unique probability  $p$ ,  $Y_2 \approx pY_1 + (1 - p)Y_3$ , where  $\approx$  indicates indifference

**Independence** requires that if  $Y_1 \geq Y_2$ , then  $pY_1 + (1 - p)Y_3 \geq pY_2 + (1 - p)Y_3$

There is good evidence that individuals in practice do not conform to the axioms of EUT (Tversky and Kahneman 1986), although it can still be argued that rational people ought to conform to the theory (Broome 1991). For example, it does not necessarily follow that a self-interested but risk averse individual would automatically prefer an alternative with the best odds, over an alternative that provided lower odds, but a certainty of at least a small win. Broome has illustrated this with the example of a gamble of 99 units for certain, with a 50:50 chance of 200 or nothing. EUT would predict the individual to choose the latter option, but the individual might prefer the choice of 99 units since this is the 'safe bet'.

However, the contribution of von Neumann and Morgenstern's theory is in the characterisation of how rational individuals *should* make decisions when faced with uncertain outcomes, where probabilities are not objectively known. They outlined axioms of choice that have become basic foundations of decision analysis. Raiffa (1968) and several others (Bell 1982, Howard 1992) have expanded this work.

The empirical findings do not violate the scientific rationality of expected utility theory *per se* since the axioms need not be descriptive to be normative (Nease 1996). Because people do not make choices that are consistent with EU theory, this does not mean they ought not to make decisions based on EU theory to maximise their utility. It is a theory built on logical mathematics, not on how people are observed to behave, so this is not a surprising finding. It is argued that EUT provides a metaphor for how people make decisions under uncertainty

(Nease *ibid.*) and as such can generate valuable insights, even if this is not how individuals actually behave all the time. What is important is that the empirical work founded on the principles of EUT recognises the normative characteristics of this theory and interprets the findings of research in light of this. No representational system can be exact or completely analogous with reality, nor can it generate a description of the world that matches it fully (Bronowski 1966). This is a theme that will be returned to in later discussions of methods of economic evaluation of health care.

### **Different intellectual pathways of economic evaluation**

The arguments presented so far have suggested that the origins of welfare economics are based on subtly different theories about the purpose of economics and the role of economics in deciding policy questions. Utilitarianism is founded on principles that have at their heart strong assumptions about the nature of human behaviour and how decisions about human welfare ought to be made. This has been contested by economists who argue that every decision rule has a value judgement and moral position on which it is based, and do not provide technical solutions to social issues.

The difference in interpretation of how economics ought to develop has led to different areas of focus for economics. This divergence in pathways in the development of economics is still evident in modern welfare economics today. They remain two rival branches within the same discipline. Sen has proposed that modern welfare economics has dual origins in both ethics and engineering and that this explains the duality of concerns and interests (Sen 1987). This mirrors the schism in intellectual approaches set out earlier. One pathway has been to take economic thought (and economic evaluation) into the area of mathematical complexity, using methods developed in operation research and statistics. The other has been to maintain some political, ethical or moral dimension of economics and economic questions. It has been argued that questions of human well-being and welfare have been marginalised and the egalitarian element of the utilitarian doctrine has been sterilised by moving from utility to physical output as the object to be maximised in modern economics (Robinson 1964).

A dominant focus in economic evaluation has been in the validity and reliability of measurement that has attempted to mirror physical sciences. Interest in the mathematics of human behaviour and psychology (which is important in later sections of this chapter) has developed within this tradition. Political economy, which has at its core a moral focus on how society ought to be, has branched into a separate discipline although the forefathers of economics saw moral philosophy, politics and economics as part of the same core discipline.

However, it could be argued that all economics has at its core a value base or moral philosophy. The split between political economy and economics is over whether the consideration of moral philosophy ought to be explicit or implicit in the analysis. The moral philosophy in economic evaluation can be gauged by whatever is being maximised by the rational individual. Sen has distinguished three types of motivation: self interest, sympathy for others and duty or commitment to moral principles (Sen 1987) Economic evaluation is usually restricted to the maximisation of self-interested individuals.

Another way in which economics has branched away has been to reject some of the central tenets of welfarism as unworkable in the real world and to dedicate energies to “the activities prohibited by the theories and axioms of the hard core of the parent discipline” (Edwards 2001), specifically the interpersonal comparisons of utility. The efforts of the “extra-welfarists” in health economics have been steered towards the search for an operational definition of health gain as the unit for its cardinal measurement, thereby bypassing the problem of the impossibility of utility measurement or comparison. By attempting to replace utility with other comparative units of health, health economists in particular have moved away from traditional welfare economics. It has been suggested that this approach has made the discipline more vulnerable to “an intellectual vacuum [that] looms large in the history of economics” (Blaug 1985).

Consequently, there is a tension between theoretical validity and practical application at the core of health economic evaluation. This theme will be returned to later in this chapter and in subsequent chapters. The next section considers approaches to economic evaluation and the extent to which they adhere to the principles of welfare economics discussed here.

These debates are reflected in the approaches that will be outlined in the following sections on approaches to economic evaluation since they about trying to understand what is “good” for society and how this can be resolved. The satisfaction of individual preferences is at the core of welfare economics, but not at the core of health economic evaluation. The question of how to value complex service that aim to satisfy people’s needs rather than to produce objective health benefits is especially relevant to the discussion of how to evaluate services such as palliative care.

### **Cost-benefit analysis (CBA)**

Cost benefit analysis is the pragmatic application of modern welfare economics that compares the overall gains to the overall losses and considers a policy to be an improvement if there are more gains overall. In order to create the greatest benefit at the

least cost as those costs and benefits are viewed by the people who experience them it is necessary to aggregate individual preferences into global preferences. There is no completely objective unit of measurement for doing this, but money can be used as a measure of subjective value (Scitovsky 1993). CBA considers the subjective social costs and consequences of interventions, and values them in the same common (usually monetary) units for comparison with all other possible uses of the same resources (Winch 1971). Decision rules of CBA can lead to an optimal level of allocation of resources if all gains and losses can be measured satisfactorily in this way. It has been argued that this approach can be applied to the evaluation of market and non-market goods (Hanneman 1984).

CBA has been described as a form of “simple democracy” as it records preferences (how much society desires something, how much they would sacrifice for it) and make recommendations on that basis (Dasgupta and Pearce 1972). Implicit in this is that individuals’ preferences should count. The foundations of these preferences and what leads to individuals’ decisions to consume (selfish wants, desire for equity, altruism) are not considered. It is well founded in rationality as a way of identifying what society prefers and is able to approximate social preferences and measure social gains and losses.

This is referred to as a ‘gold standard’ approach in economic evaluation (Mishan 1988) as all the costs and all the benefits (the gainers and the losers in any intervention with economic consequences) are defined. CBA takes into account all the social costs and benefits of a policy (or economic change) rather than making *a priori* decisions about which costs and benefits are worth measuring. It has been described as “the most powerful aid to public decision-makers” as it attempts to replicate market processes and produces results that can be compared with private sector appraisal techniques (Hutton 1992).

A profusion of technical and measurement difficulties surround the practice of CBA (Ackerman 1997). One particular problem with CBA techniques in real world evaluation is they ignore distributional equity: who gains, who loses, and the final distribution of income. For example if a change makes people on low incomes worse off by £100 and the rich better off by £250, it still produces excess gain of £150 for society as a whole. CBA would favour such a change if distributional considerations were not taken into account. As Hicks-Kaldor type compensations between gainers and losers in an economic change are only hypothetical compensations; a Pareto improvement is consistent with many people actually being made worse off. This is unlikely to receive support from the public, even after the system of taxation and subsidies to bring about a redistribution of wealth and increase welfare more evenly (Mishan 1988).

There has been some suggestion of weighting gains and losses to reflect equity considerations, but Mishan has argued forcefully that this leaves CBA exposed to political lobbying (and no better than it) since it opens the possibility of making the results seem more favourable for particular groups (Mishan 1988). He argues that public policy changes should meet the tests of the political process independently of CBA and leave the method of assessing gains and losses as politically neutral as possible (within the constraints of the explicit behavioural assumptions inherent in the methodology).

Critics of CBA have questioned the psychological realism of hypothetical valuation without giving individuals any real context to the evaluation problem (Frank 1989). The notion that the social optimum can be reached by satisfying private preferences has also been challenged (Sagoff 1994), as has the idea that non-market goods such as health can be “commodified” (Anderson 1993). This last point questions the assumption that value can be determined by the same method as market transactions for private goods and that they are therefore substitutable with other commodity bundles that have an equivalent value. The approach assumes that market norms and private preferences should shape policy. Proponents have argued that the validity of CBA is precisely because it does value non-market goods and is oriented towards valuing all costs and benefits regardless of whether they enter financial calculations (Campen 1986).

Mishan (1988) argued that there might be grades or stages of CBA and society can decide how far it is willing to accept the premises of Pareto’s criteria for welfare improvement. Society must decide how much it is willing to trade Pareto optima in resource allocation for other values it holds important, such as distributional justice. Practitioners of economic evaluation whose analyses feed into the decision-making processes must decide which approach to adopt. This could be the utilitarian model of welfare (where the total amount of utility is the final goal), or they could take account of the distribution of costs and benefits among individuals in society. This has direct relevance to the current debates in health economics on the equity-efficiency trade-off discussed further on.

Yet Mishan was optimistic about a reduced role for CBA in public policy analysis. He saw Pareto optimality as a “guide not a religion”. CBA can provide a sum of valuations given by the public, but no warrant for asserting the ranks or preferences arriving from these valuations that may need to take other valuations into account. This he saw as best left to the politicians and the democratic process. This left CBA with a “more modest but useful (if still overrated) task” (Mishan 1988).

Concepts of liberty and coercion and the role of government in determining what is 'good' for society have added to the complexity of the arguments about how to determine "value" but they have not had a major impact on the underlying core of welfare economics in economic evaluation. These issues are not generally considered within the economic evaluation framework that has tended to focus on technical concepts of efficiency. Social choice theory has had some impact on economic evaluation, especially in the area of poverty reduction and economic development, (Sen 1982) but this has tended to be at the macroeconomic rather than microeconomic intervention level. The lack of a clear decision rule to replace Pareto optimality as the yardstick by which to measure welfare change has meant that it has not had such an impact on the microeconomic analysis.

There are problems with applying CBA to health care that have led to the development of other methods of evaluation becoming dominant in health economic evaluation. These problems and the reason for the dominance of alternative methods is explored in the next section, alongside the debates about the extent to which these alternative methods have retained or moved away from the central tenets of welfare economics. All methods of economic evaluation make a trade-off between adherence to theoretical principles of Pareto Optimality (and application of the logic and decision-rules arising from the theory), and the practical need for information (without logical decision rules). The problem that this compromise raises, and the implications for what can be measured and valued using these other methods, is the focus of the next section.

### **Summary of the welfare economic arguments relevant to the thesis**

This review of the origins of welfare economics and its practical application in CBA is presented to consider the fundamental assumptions that underlie economic evaluation. It has been necessary to go back this far in order to have a basis for understanding where economic evaluation in health care comes from, the extent that it adheres to this underlying theory and the problems that arise from moving away from these underlying principles. This is not to say that these principles should always be strictly adhered to, or that economic evaluation methods that have moved away from these principles without good reason. But it recognises that all method of economic evaluation that are not strict cost-benefit analyses make trade-offs between theoretical validity and practical application. The argument is that departure from theory matters in some contexts more than in others. In the evaluation of complex areas of health and social care, where the relationship between health gain and the value of that health gain is not straightforward, approaches that adopt the health gain approach may not capture the value of these services to the individuals who use them, and to society more widely.

## **SECTION 2. METHODS OF ECONOMIC EVALUATION OF HEALTH CARE**

The first evaluations of human life concentrated on valuing the changes in quality and quantity of labour (the human capital approach) (Mooney 1977). This approach took as its starting point the notion that investment in education, training, and health care would have an effect on labour and national wealth. It did not consider the outcomes of health care beyond the benefits to the labour market. Neither were consumer preferences included in this valuation. In this simple framework, society has no input into the valuation either of human life or on how public money should be used to improve life. The narrowness of focus has meant that this approach did not have much application after the 1960s, and 1970s (Mooney *ibid.*)

Dreze (1962) has been identified as the first to consider the valuation of life along lines that link it theoretically with welfare economics, with considerations of consumer sovereignty and with reference to individuals' preferences. He proposed that if the costs incurred by an individual to avoid the risk of premature death could be freely determined by an individual (unconstrained by coercion or incomplete knowledge), then the monetary value would reflect society's preferences. Schelling (1968) emphasised the need to reflect the valuation of one's own life rather than someone else's (therefore consistent with the individual consumer sovereignty approach). Much of the early work focused on saving life rather than the value of life itself, which represented a move away from the human capital approach.

The insurance-based approach might also be interpreted as reflecting a completely different quality in the valuation of human life (fear of death, psychic attributes of life). However, trying to establish the value of life by assessing the value of risk of death is flawed since an individual's life is not replaceable if it ends. An individual will never be indifferent between the risk of death and a monetary value (there is no "claims" market for one's own life) (Jones-Lee 1976).

Mishan was highly critical of these early attempts to value life, and endeavoured to put human valuation firmly into the methodological paradigm of cost-benefit analysis, guided by Pareto optimality (or potential Pareto improvement) and compensating losers in a policy change (Mishan 1988).

### **Advantage of CBA in health care evaluation**

CBA allows all the important influences on demand for a health service and future welfare from that service as perceived by the individual to be taken into account. For any description of reality, or a health scenario they might face, individuals can value the consequences of



that scenario by taking into consideration any number of variables, both known and unknown (Pearce and Nash 1981). This includes attitudes to risk, attitudes to future health, and views on equity (benefits to other individuals) (Mooney 1977).

The main theoretical advantage of CBA over other evaluation methods in the health sector is that health and utility derived from health is only one argument in the social welfare function. It starts from the assumption that the only necessary data are orderings or subjective valuation of individual members of society. No other principles of the general good are needed. Actual Pareto improvements are extremely rare, so the relevant question is whether a policy or intervention brings about a potential Pareto improvement. If, in aggregate, the net gain is greater than the net loss, then there is net social benefit.

### **Problems of applying CBA to health care**

There are specific problems in applying CBA to health care. In the 1970s, Mooney set out the arguments for why CBA had not had a significant impact on decision-making in health care (Mooney 1977). He argued that there had been a lack of theoretical progress since the rejection of the human capital approach; that the power of the medical establishment had not been supportive of introducing economic evaluation; the decision-making process was diffuse in health care; the outputs of health care were also diffuse; and there was a lack of information of effectiveness on treatments. Mooney also argued that only a small proportion of resource actually saved lives as a primary outcome, while the rest was used to decrease morbidity, the measurement of which is fraught with difficulties. Mooney also put the case clearly that the valuation of human life is not a numerical sum since there is no simple calculus for comparing the cost of a health intervention with the benefits of saving or improving a human life (Mooney 1977). There is no "market price" for life to determine its monetary value. All these problems, he argued three decades ago, have contributed to a lack of demand for CBA in health care.

It has been argued however that CBA is a useful conceptual idea where markets did not exist, as addressed by Mooney:

*"All we require to establish that something has a positive value for which a monetary value could exist is that individuals in society would be prepared to pay a positive monetary sum for that 'something'" (Mooney 1977, pp56).*

If market price and consumer surplus indicate the value of a commodity, then, Mooney argued, consistency dictates that methods of evaluation should be attempted in order to

establish the monetary value of non-market goods. The central assumption of this line of argument however, is that the axioms of the market also hold for non-market goods. Consumers are seen as sovereign, fully informed individuals who know how to satisfy their own welfare needs, and their needs can be satisfied through consumption. If this is the case, then a monetary valuation (what individuals would be prepared to pay on the open market) would reflect the value of these commodities to individuals. These conditions may not be achieved but it is important to remember that these are the underlying assumptions of the approach. By the beginning of the 21<sup>st</sup> century, some of the issues had been addressed to some extent, but barriers, both practical and theoretical, to implementation of CBA have to a large extent remained (Dolan and Edlin 2002). The measurement of morbidity and quality of life remains a major challenge for economic evaluation. There is also a problem in assuming that individuals are welfare maximisers. Advertising, and differences in people's tastes mean that people can become divorced from the value of their consumption. As a consequence, their consumption of some goods bears no relation to their anticipated welfare.

The theoretical challenges of CBA have been expanded since Mooney's analysis. Dowie has questioned whether the CBA approach is appropriate for the evaluation of publicly funded interventions (Dowie 2002). He argued that the valuation of private goods and public goods is different, and has questioned whether values for private consumption should determine how publicly funded goods are distributed. This is somewhat similar to Sen's argument that only some values should count in social choices (Sen 1970b). Sen's approach is based on the assumption that the problem of collective choice cannot be satisfactorily discussed within the confines of economics since Pareto Optimality and liberalism are mutually exclusive in some contexts. But these arguments are not relevant to CBA alone, and could relate to all methods of valuing the outcomes of health care. In one of the few studies that have looked at the effect of the perspective on individuals' valuation of benefits, Dolan and colleagues argued that whether the respondent considers his/her own (private) utility, or that of society had no discernible effect on their views (Dolan and Cookson 2000).

### **Rejection to CBA and adoption of cost-effectiveness analysis (CEA)**

The most straightforward and the most common form of economic evaluation of health care interventions is CEA. Achieving the same results more cheaply represents a "classic cost-effectiveness result" (Donaldson 2002). Cost-effectiveness analysis addresses the problem of health outcome measurement by expressing benefits in physical output, rather than metaphysical constructs such as utility. These can be improvements in physical or functional status (Palmer et al 1999). The relationship between physical output and welfare is implied

or understood to be direct and the same for all individuals. The concern is for economic efficiency only without explicitly addressing issues of justice and fairness of the conclusions of research. Increasing efficiency implies a change that is a potential Pareto improvement regardless of who potentially gains and loses. (Kawachi 1989).

CEA is a method of comparing programmes interventions within a single disease or where the interventions are directed towards the same health objective. Since benefits are not measured in comparable units, the efficiency of programmes can only be gauged within one health objective to achieve technical efficiency. Cost-effectiveness analysis adopts a more limited view of the benefits of health care interventions. For this reason it does not investigate whether a given health care intervention represents a Pareto improvement or a potential Pareto improvement. Thus it is clear that an efficient allocation of resources, in line with the principles of welfare economics, cannot be determined or achieved following the rules of CEA.

However, the attraction of the CEA approach is that the measurement of simple, practical concepts of health outcomes (such as cases detected or patients treated) avoids the problem of evaluating whether individuals are the best judges of their own welfare, and of measuring utility. The problem is that not all aspects of utility or welfare may be incorporated into the analysis, and preferences for the intervention (as opposed to other uses of the same resources) cannot be considered.

This approach has taken economic evaluation in health care away from its roots in welfare economic theory. Final or intermediate outcomes can be objectively measured if clearly defined, and can be compared across individuals. It assumes that one person's treatment produces the same utility/happiness/welfare as any others, which in many contexts is an acceptable, plausible assumption. The chosen measures of effectiveness are straightforward to conceptualise and the problem of *valuing* benefits is avoided. Recent work by Dolan and colleagues have rejected the idea that CEA and CBA can be theoretically linked through the identification of a constant willingness-to-pay for a health outcome and that these approaches should be seen as distinct and not related to each other through some underlying scale of utility (Dolan and Edlin 2002).

CEA has emerged as a dominant paradigm in economic evaluation in health care partly because of these important and practical properties (objective measurement and assumption of equal utility from the outcome across individuals) (Hutton 1992). Another reason for the strong support for CEA studies in health economics is the dominance of economic

evaluation alongside clinical effectiveness research. In these studies, health outcomes are the endpoints of interest. The trend towards CEA avoids the very complex practical problem of the multidimensionality of benefit of health care and the value-laden objectives of health care (Brouwer and Koopmanschap 2000).

Economic evaluation guidelines have identified the important criteria for CEA: unambiguous, simplified objectives based on existing clinical evidence (or obtained concurrently in clinical trials) and where, if possible, there is only one major dimension of benefit. The evidence of effectiveness, then, is “primarily an epidemiological problem” (Drummond 1997, pp98). The ‘problem’ for CEA is establishing the relevance of the results of clinical trials (which usually measure efficacy rather than effectiveness) (Gray et al 1997; Torgerson and Byford 2002). Drummond suggested that results can be adjusted to take account of this by using sensitivity analysis to test whether different assumptions change the final result (Drummond, O'Brien et al 1997). Much of the economic evaluation work now undertaken in health care is based on this approach and comprehensive bibliographies of health care economic evaluations in the 1990s (Elixhauser 1993, Elixhauser 1998) have shown significant growth in the literature. These reviews show that the literature is dominated by studies of specific prospective new interventions compared with current practice.

The value of the CEA approach are clear: it provides a simple and straightforward approach to comparing ‘simple’ interventions with the same health objectives (Brazier and Dixon 1995). “Simple” means the endpoints are well described and measurable, the intermediate endpoints relate to final outcomes in clear and direct ways, and the intervention is approximately the same for all patients/users (Drummond et al 1997).

One of the limitations of this approach is that it is only a partial form of analysis. Therefore it may fail to identify current misallocation of resources by focussing on the evaluation of technologies or strategies in a narrow field of enquiry. Thus, it cannot address questions of allocative efficiency, that is, whether the extra cost is worth the extra benefits of a change in policy (Donaldson 2002). The results of most if not all CEAs are so context specific that they cannot be used to inform the wider policy debate (Murray et al 2000). While this is also a criticism for all types of economic evaluation, it is particularly relevant to CEA where outcomes are narrowly defined that they may have no purpose outside a highly specified clinical area.

Clearly, there are many evaluation problems where these conditions or criteria are not met. Where interventions have an impact on different aspects of welfare, for example on quality of

life rather than life expectancy for example, a methodology is required that can compare different types of outcomes.

### **Cost-utility analysis (CUA)**

The dominant focus of CEA on only one outcome of interest to the exclusion of other important outcomes has led to an interest in more sophisticated measurement tools to capture multiple dimensions of benefit. There are two reasons for creating a more sophisticated tool for measuring effectiveness than considering only “natural units” of outcome. The first is to combine more than one dimension of benefit into a single unit, and second, to address the *value* of the utility (i.e. the welfare derived) from an intervention to the individuals who receive it. This second reason is relevant since it is an important source of contention in the literature

Researchers have been attempting to develop more sophisticated indicators of morbidity and mortality that represent quality as well as quantity of life in one combined index. This index should be capable of reflecting different health experiences of the population. CUA provides a way of trading length of life with valuations of quality of life and so provides a way of comparing outcomes from different types of health interventions (Gerard 1999). The standard form of measurement is the quality adjusted life year (QALY)<sup>2</sup>. It captures gains from both the prolongation and quality of life in a single measure, and can incorporate the value that people place on different health outcomes, or their preference for different health states. Programmes with the lowest cost per QALY would be recommended for prioritisation with the aim of maximising health gain under budget constraints. It is therefore argued to be a superior way of comparing interventions across a diverse range of health care (Gold, Weinstein et al 1996). QALY analysis measures health gain in specific domains of health status. The approach also incorporates some notion of value for specific health states. These values can be derived either from people who have experience of a health state (i.e. within an empirical study) or from values derived from the general population, thereby reflecting society's value. Contrasting approaches represent different forms of CUA analysis that make different assumptions about what should be measured and included in the analysis. The different approaches are reviewed in this section.

The literature on QALYs is more than thirty years old. Despite differences in the theoretical basis of different forms of QALYs, there are also important similarities. All the approaches

---

<sup>2</sup> Disease adjusted life years (DALYs) are not critiqued separately, as the same debates can be applied since both are ways of quantifying life years adjusted for differences in quality of life (Arnesen and Nord 1999).

assume a full year of health is scored as 1 year of life and less than optimal health is scored less than one. The assumption of additivity means that two years of life scored at 0.5 is equivalent to a full year of health life. This means individuals will be indifferent between these choices and will not trade between them to reach a higher level of utility.

The increased interest in health outcomes rather than processes or throughput has encouraged this investigation (Kind, Dolan et al 1998). Cost-utility analysis (CUA) is seen as more compatible with the decision-making environment, especially where quality of life is the important outcome (Torrance 1986). This class of studies are also designed to inform decisions about whether an intervention service should be implemented.

An extensive review published in 1992 counted 51 such studies, a large proportion of which were found to be deficient methodologically (Gerard 1992). Nevertheless, the CUA framework has been suggested as the reference framework for health care economic evaluation (Weinstein et al 1996). A later review of the quality of 228 published CUA studies showed that the field had increased greatly between the mid 1970s and 1997 with a wide variation in methods of eliciting preference weights. (Neumann et al 2000).

### **Preference-based versus health-state based QALYs**

There is a continuing difference of opinion about whether the role of the health service is to improve people's health as defined objectively or to improve welfare, defined subjectively (Nord 2001, Johannesson 2001, Williams 2001). There are also contrasting approaches to deriving QALYs based on how outcomes are defined: one approach measures outcomes as the utility or value society places on health states (utility-based approach). The other rejects many of the fundamental principles of welfare economics, especially its focus on utility-based ideas of welfare, and defines the outcome as a quantity of health (the extra-welfarist approach). Whether or not QALYs can or should be based on individual's preferences and values has led to different approaches to QALY estimation. There have been debates through the 1980s and 1990s on the best methods of deriving measures to reflect the true utility from a health intervention. Much of the debate has been around the economic theory behind the derivation of utility.

### **Preference-based approaches to conceptualising and measuring dimensions of health-related quality of life**

There have been attempts in the economics literature to develop generic scales that reflect health-related quality of life using generic scales to measure cross-diagnosis outcomes. An early example of this was the Rosser Distress and Disability Matrix (Rosser and Kind 1978)

where utility scores were devised for specific health state by asking a sample of 70 people to score each health state on a scale of 0 to 1. This score was used in the first derivation of the quality-adjusted life year weights. Later examples used in the health economic evaluation literature have been the Quality of Well-Being Scale (QWB) (Kaplan et al 1993), the Health Utility Index which has now been produced in three versions, the HUI 1, 2 and 3 (Torrance and Feeny 1989, Feeny, David et al 1996, Furlong, Feeny 2001a), and the 15D scale (Sintonen 2001).

In a recent review of methods, the Rosser scale and 15D scale were considered to be inferior because the valuation of health states as not been derived from choice-based techniques for establishing the utility values and therefore not based on von Neumann-Morgenstern expected utility theory. The Rosser scale has also been shown to be less sensitive to changes in health states than the EQ-5D (Brazier et al 1999).

There have been three main methods of establishing utilities for health states: standard gamble, time trade-off and person trade-off (Nord 1995; Dolan 2000). Each method reflects different concerns about the way in which utilities ought to be measured. There are strong arguments about which method (or series of methods) most accurately reflects the decision-making context that patients have to face in reality. Standard gamble (SG) is based on well grounded theory from operational research on behaviour under risk (von Neumann-Morgenstern expected utility theory) but the hypothetical choice may be highly unrealistic and respondents may have difficulty in comparing probabilities of compete health or death versus certainty of limited health. The HUI has adopted the SG method of eliciting health state values.

Time trade-off encapsulates choice under uncertainty but rejects the underlying von Neumann-Morgenstern expected utility theory and the notion of risk-taking. Person-trade-off reflects more accurately respondents' choices about social rather than individual utility (Nord 1995) but the approach has been criticised as being too high a cognitive burden on respondents who must weigh up a large number of variables, such as severity of ill health before treatment, after treatment, overall health gain and number of people who benefit (Dolan 1999). The EQ-5D has adopted this approach in valuing health states.

The SF-6D, a preference version of the widely used SF-36 has been developed (Brazier, Usherwood et al 1998). The measure was developed as a way of deriving a preference-based utility measure from the widely used (non-preference based) SF-36 instrument.

The recommendations of a Health Technology Appraisal that reviewed health status measures in economic evaluation came out strongly in favour of using either the EQ-5D or the HUI (Brazier et al 1999).

A recent article has reported that, rather than attempt to replace EQ-5D, this approach was developed to make use of SF-36 data that had already been collected in clinical trials and “where there is no other means of estimating the preference-based health values for generating QALYs” (Brazier et al 2002. *ibid.* pp289). The authors suggest that the SF-6D may have greater sensitivity than the EQ-5D since it is derived from a much larger descriptive system (36 domains from the SF-36), but this has not yet been tested empirically. The authors also report that regression coefficients reported for the extremes of the scale did not produce statistically significant estimates, suggesting that there may still be some inconsistencies in the values derived for the instrument.

### **Estimating QALYs from health-related quality of life scales**

Adopting the approach developed for the EQ-5D, the derivation of QALYs involves asking individuals to rank their health state along a series of simple health dimensions. This is a method of deriving utility weight estimates for specific health states. The dimensions are: mobility, self-care, usual activities, pain, and anxiety/depression. For each dimension, there are three levels an individual can classify their state of health in: having no problems in this dimension, some problems, or having extreme problems. From this scheme, 243 possible health states can be described. Each state of health is given a weighting that indicates how near or far the health state is perceived to be to death (0) or perfect health on a continuum. In theory all states of health from those close to death (or worse than death) to perfect health can be described in a combination of these levels. Time in these health states is multiplied by this quality of life weight to give a composite score. A QALY represents a full year in a health state adjusted for the value of quality of life in that state. All QALYs are assumed to be perfectly divisible, and individuals to have perfectly smooth indifference curves for quality and quantity of life. For example, an individual is assumed to be indifferent between 4 years in perfect health and 16 years in a health state with a weight of 0.25.

One group of economists have argued that QALYs should be based on time trade-off methods (TTO) as the true ‘gold standard’ methods (Torrance 1986; Mooney and Olsen 1991) and that TTO poses the ‘right’ questions for QALY derivation. TTO is proposed as an empirical substitute to standard gamble techniques that provide similar results but is easier to administer. This means therefore that it can mirror expected utility theory even if it is not directly derived from it. Another group has rejected this method, and developed another tool



(Healthy Years Equivalents, or HYE) derived directly from standard gamble techniques. They argue that this method more accurately the true utility derived from an intervention by incorporating respondents' attitude to risk and uncertainty about future health needs and the effectiveness of health interventions which TTO does not (Gafni et al 1993; Gafni and Birch 1995, Gafni 1997, Gafni and Birch 1997). The supporters of the QALY derived from the TTO approach criticised this argument, saying that HYE derived using SG techniques were essentially measuring the same values (Loomes 1995), and that criticisms for or against one methodology could equally be applied to another (Buckingham 1993, Culyer and Wagstaff 1993).

The outcome of this debate is that there are important differences between economists in how they see the measurement of the benefits of health outcomes. Utilities represent the subjective preferences of individuals. As measures, utility scores are meant to provide quantitative estimates of individual preferences. In the context of health care decision making, utilities reflect individual preferences for particular health states. Therefore, there is a clear distinction between utility or preference-based measures of health status, and non-preference based or descriptive measures. The utility-based measures differ from each other in whether individuals' valuations of benefit incorporates risk and uncertainty; whether the instrument which measures utility is too onerous on respondents and whether the decision-making reflects the choices facing individuals in the real world.

Finally, the assumption that QALYs have cardinal properties could be said to either resolve the problem of interpersonal comparisons of utility (ICU), or simply to side-step it. The approach assumes that QALYs have cardinal properties, in that they can be added together to produce a sum of total social welfare. This can be interpreted as a fundamental flaw in terms of their adherence to the Paretian framework of welfare economics or a clever avoidance of this complex and irresolvable problem.

### **Utility-based QALYs**

The history of utility theory and its application to health outcomes has been reviewed by Torrence and Feeny (1989). This approach to estimating benefit brings economic evaluation closer to the original principles of economic evaluation than CEA as it establishes the benefits to individuals of an economic change (an intervention) as measured by their own valuation of that benefit (Mooney and Olsen 1991).

For utility-based measures of QALYs, outcomes are interpreted in terms of personal valuation (reflecting individuals' preferences) rather than amount of well life (Culyer and

Wagstaff 1993; Nord 1994). Since it is not possible to observe directly the revealed choices of individuals for different health states, this is elicited by asking people to imagine themselves in different health states. They are asked to assess their hypothetical willingness to trade their wealth (or other domains of welfare) in order to change to a better health state, or to be compensated for remaining in a bad health state (Nord *ibid.*). By asking these hypothetical questions, it is possible to establish a person's expected utility (or value) from that health state.

Methods have been developed to measure the utilities (more accurately called QALY weights) of particular health states. The standard gamble method asks respondents to make a choice between the certainty of a long-term condition for a specific number of years, or a gamble between the probability of complete recovery and a small chance of death (Johannesson 1996). The respondent chooses between the two alternatives: certainty and a gamble. The probabilities in the gamble alternative are altered until the respondent is indifferent between the first (certain) choice and the gamble. The probability where the respondent is indifferent is interpreted directly as the utility of the first alternative.

The time trade-off method for estimating 'utilities' for health states is an approximation of the standard gamble approach that overcomes some of the problems of dealing with probabilities. It involves the respondent making a choice between a chronic condition for a number of years or a treatment that will result in perfect health but shorten life span. The shorter time period is altered until the respondent is indifferent between the choices. The utility is derived from the ratio of time in perfect health over time in poor health (Johannesson 1996).

Expected utility theory as the theory of choice under uncertainty has provided the theoretical basis to cost-utility analysis. The validity of expected utility theory and its underlying axioms have come under increased criticism as suggested earlier in this chapter. Empirical studies in the health field have shown that individuals will systematically violate these axioms (Oliver 2002) and that these violations cannot be explained away by anomalies or errors or ill-constructed thought by respondents. These axioms have also been challenged by authors of empirical studies as too restrictive and not reflecting reality (Sackett and Torrance 1978; McNeil and Stones 1986). McNeil found in his empirical work that older individuals were only willing to trade longevity for quality of life when the length of time in less perfect health is more than five years. Sackett and Torrance found that values which respondents place on health states varied significantly with the length of time in that health state (Sackett and Torrance *ibid.*). Also, there must be an independence and constancy in risk attitude to

survival duration. However they found strong evidence that individuals exhibit both risk taking and risk aversion, that is, no consistency.

The practical and theoretical problems remain and there are strong doubts about the possibility of deriving valid and reliable estimates of utility weights based on expected utility theory. (Loomes and McKenzie 1989, Dolan 1999). Patients' preferences may not fit the model proposed from expected utility theory. Some researchers (economists and non-economists) have questioned the predictive usefulness of EUT as a model for rational behaviour under uncertainty leading to the development of alternative theories of behaviour such as prospect theory (Verhoef et al 1994, Treadwell and Lenert 1999); regret theory (Loomes and Sugden 1982, Mooney 1989, Smith 1996) and disappointment theory (Bell 1982).

### **Rejection of utility-based measures of benefit – Extra Welfarist approaches to economic evaluation**

This approach has developed out of a frustration with the practical problems of measuring the utility from health interventions and the rejection of the fundamental axioms of welfare economics. The analysis can embrace whatever maximands the customers of research may give, or may be inferred by diligent enquiry by the analyst to be relevant.

The necessity of adhering to welfare economic principles has been interpreted as being either not useful or too unwieldy for real-time economic problems. This has led some analysts to reject welfarism altogether (Culyer and Wagstaff 1993, Culyer 1997). Since there is no scientific way of resolving the problems such as the impossibility of interpersonal comparisons of utility, one proposal has been to simply ignore this issue and instead consider units of health as the endpoint of health interventions (Williams 1996). The simplified argument is that other factors apart from welfare maximisation influence individuals' consumption and preferences, and that these non-health maximising preferences should not be used to determine health policy (Mooney 1997). Furthermore, health care planning incorporates normative, political judgements about health care and individual preferences cannot override these judgements in a publicly funded system. Frustrations have arisen in health economics from trying to adhere to the criteria that must hold for Pareto optimality. This has led to an abandonment of these principles as unwieldy and unhelpful. (ibid.1997).

The foundation of this argument is that society has already decided (through the political process) that health for its own sake is a desired outcome of health services. For this

reason, it is acceptable to measure the health outcomes of an intervention without considering the utility derived from them. The important outcomes therefore are commodities themselves: these are commodities containing the specific characteristics to improve health care.

The objective in this analytical framework is health maximisation rather than utility. Distributional concerns (who gains, who loses) and equity can be addressed by additional weights for different groups of recipients of health care (for example children, or the elderly) which reflect the values of society rather than the value individuals place on their own welfare (Bleichrodt 1997). Utility is therefore derived directly from a health state: the assumption is that two individuals cannot be in the same health state and derive a different benefit from that health state. The emphasis is placed firmly on the practical application of cost-utility analysis. The economist is the consultant who contributes to the decision-making process by presenting policy choice options and spelling out the consequences of these options for the policy customer. The distinction between Paretian approach to CBA and this decision-making approach is the focuses on addressing the needs of decision-makers and making the evaluation consistent with their objectives. This is the important distinction between welfarist and extra-welfarist approaches (Sugden and Williams 1978).

Sen introduced the idea of 'basic capabilities'. The argument focused on the capability to function, what a person can do or can be. It questioned the more standard emphasis on maximising wealth and utility and suggested replacing the immeasurable concept of welfare with more finite objectives (Sen 1985). Within this framework, economic analysis could adopt any objectives given by the customer (decision-maker). The task of the economist was therefore to find out the relevant issues for the client. Sen argued that this agenda provided a methodology that could be more thoroughly researched than the welfarist approach (Sen *ibid.*).

Culyer has built on Sen's work by distinguishing between things and the characteristics of things. His evaluation framework is based on the idea that the important outcomes of health technology are not utilities, but objective measurable commodities (Culyer 1990). This avoids the metaphysical question of whether the important outcome of a health care intervention is the effect on health or on utility. He argues that since overall welfare is beyond the remit of the health service, so utility should be beyond the remit of health economic evaluations, with the objective of health maximisation rather than unknowable and immeasurable latent concepts of utility.

Williams has argued that there is a whole set of ambiguities buried in the utility-based approach to measuring outcomes, that go beyond the objective and verifiable facts of any comparative decision that requires comparative judgements between individuals and over time (Williams 1992). They also concern who should be making the judgements and whether these judgements should influence policy.

Furthermore, because the Paretian framework does not allow individuals to make interpersonal comparisons of utility, then it was argued that the welfare economic approach to economic evaluation is untenable and ought to be abandoned (Williams 2000). Further, the measurement of welfare in terms of money (or willingness to pay for health care) does not avoid this problem as social welfare maximisation is defined as the point at which no change exists where the monetary value to gainers is greater than the monetary value to losers. This implies comparing values between gainers and losers. This Williams called "ICU by the back door." (Williams *ibid.*) He argued that Arrow's theorem could be avoided if judgements are made explicit, and it was accepted that there is no scientific solution to this problem.

This position has had some support in the psychology literature. Arrow assumed that the metric underlying utility was not meaningful and not standardised across individuals. However, later psychometric evidence now suggests that preferences can be measured using scales that have meaningful interval scales or ratio properties (Kaplan 1993, Brazier and Deverill 1999). If QALYs are assumed to be a cardinal measures of health, then this also side-steps Arrow's theorem as such units of health can be measured, compared across individuals and aggregated (Kaplan *ibid.*).

Critics of this method argue that the measurement of QALYs as a health status index does not stem directly from the individual's utility function and thus only partly reflect the individual's true preferences. This might lead to the choice of non-preferred alternative due to the misrepresentation of the individual's preferences (Mehrez and Gafni 1992). Fundamentally, the extra welfarist approach still requires cardinal measurement and interpersonal comparisons of utility that is forbidden by the axioms of welfare economics. QALYs, however they are derived, have to ignore these axioms in order to address the practical questions that policy-makers seek guidance on.

An important difference between welfarist and extra-welfarist approaches therefore is that extra-welfarists make health an objectively measurable commodity rather than a subjective experience. In other words, extra welfare economics assumes that the health of an individual

is a knowable, testable, finite and physical fact (Hurley 1998). The progress of extra welfarism is not a fundamental challenge to the welfare economic foundations of QALY estimation. It should be seen more as a necessary modification of theory, developed as a response to the needs of decision-makers who must weigh up the costs and benefits of changes in policies. (Edwards 2001).

### **Measurement issues in calculating QALYs**

For both the utility based and health based QALY estimation methods, the derivation of the weighting attached to QALYs is not the only methodological issue. The methods of estimating QALYs also have to be designed to be sensitive and specific enough to reflect real changes in quality of life experienced by an individual. For example, the EQ-5D instrument calculates the health status of an individual on five domains of health related quality of life (EuroQol Group 1990). Each domain categorises an individual on one of three levels (no problems, some problems or extreme problems). To detect change in (health-related) quality of life, an individual would need to move from either have 'extreme problems' to 'some problems', or from 'some problems' to 'no problems' on at least one of the dimensions of health to show a difference in quality of life weighting before and after an intervention. For interventions that have an effect on morbidity only this approach can be problematic.

However large a change in health state, it is the quantity of life that dominates the calculation of QALYs. A large change in quality of life (say from a weighting of 0.2 to 0.8) is equivalent of 0.6 of a year of full life. Therefore a comparison between two interventions, one of which brings about a large change in quality of life (0.6 QALYs) but no increased length of life will be dominated by another intervention that brings about increased length of life by one year of full health (1.0 QALY).

Second, there may be interventions where changes in quality of life are not captured by the survey instrument. The EQ-5D questionnaire requires changes in quality of life to be relatively large (i.e. moving from extreme to moderate problems, or moderate problems to no problems at all). For people with long term illnesses or disabilities, where they would not expected to regain full health after a specific intervention, an intervention may still leave an individual with some problems. However, the range or number of problems may have been reduced in ways that are important and valuable to that individual (Donaldson et al 1988). Having fewer problems as a result of an intervention but still living with many others might still register on the scale as 'having some problems' and so this change that is important to the individual, does not register in a change in quality of life. They would still have to tick the

box that recorded “some problems”, indicating that their health state had not changed in that domain. A small but significant change may be of great value to the individual.

This is an important consideration for the evaluation of complex services, especially those for people with long-term but not life threatening illness, or people with life limiting conditions. In neither of these groups of people would it be expected that an intervention would increase their life expectancy. Furthermore individuals who access these services may not expect to reach full health in the future and expect to live with some problems however good their health and social care is. This issue is raised again further on.

It is important to make the distinction here between the measurement of health related quality of life (HRQoL) and quality of life (QoL) in general. EQ-5D is an instrument to measure HRQoL. The term HRQoL distinguishes the specific health determinants of quality of life from the more general determinants of quality of life (income, status, family circumstances, job security, food security etc). An instrument which measures QoL may be less sensitive to health care issues as defined narrowly, but may be more sensitive or more appropriate when considering interventions which aim to address wider human wants or needs (Patrick and Deyo 1989). This is discussed in more depth further on.

An essential component of the decision about which type of instrument would be most appropriate for a given research context is to identify what would be considered an important change within patients or difference across patients in scores. A theme that runs through this thesis is that what is seen as a relevant change in health related quality of life to the clinician or policy-maker may not be valued the same way by the person experiencing the change. The extent to which a change has an effect on a person's overall quality of life will change from context to context. This is an argument for adopting the widest possible perspective in health care evaluation.

### **QALYs and fairness**

In health economic evaluation, there has been particular focus on whether societal concerns for fairness can be re-incorporated into economic analysis. Equity is seen as a fundamental aspect of health care, in the democratic tradition that no one group in society is more deserving than another *per se*. However, society may value (or choose to explicitly favour) the health of some members of society more than others. The problem is whether these values can be reflected within QALY analysis or whether these should be considered as separate from it.

The constructors of the QALY approach have assumed that societal value is the unweighted sum of individuals' health benefits with no regard for the distributional consequences. Moreover, a gain of one QALY for person A is strictly the same as for person B (Nord 1994). This is fair in the sense that it is democratic (no group of individuals is more deserving of QALYs than any other), but the approach also raises problems of equity. Some interventions produce more QALYs for some groups than others. The argument is that QALYs discriminate against the elderly (who have fewer years of life to live) and the long term ill (who are not expected to return to full health) (Tsuchiya 2000). While the adherents to the QALY approach quite openly admit that the approach should only be considered as a decision-aid rather than decision-maker because of these problems (Loomes and McKenzie 1989), this has been a forceful critique of the method.

Arguments that QALYs violates societal concerns for fairness have been raised in the development of QALY methods (Wagstaff 1991, Johannesson and Gerdtham 1996, Nord 1999, Johannesson 2001). Early in the development of QALYs, there were proposals to attach weights to QALYs to reflect distributive concerns (Williams 1988, Mooney 1989, Wagstaff 1991). There have been unease expressed in the literature about the ambiguity of these weights and how they can be measured and incorporated into QALY analysis. However, it is clear that individuals care deeply about equity even if it is to their detriment and overemphasis of efficiency over equity considerations may be at odds with society's wishes (Hurley 1998). A recent study has reviewed the equity problems and debates around whose values should count in valuing quality of life, those who experience the health state or whether the general public should be asked to make an *ex ante* valuations when they have not experienced the state of ill health described (Schwappach 2002).

There have been a number of attempts to reflect equity considerations including the equity adjusted year of life saved (Lindholm, Rosen et al 1998), the 'fair innings argument' (Williams 1997) and incorporating caring externalities into the calculus (Culyer 2001). But there has emerged no clear unified theory for doing this (Hurley 1998). Wagstaff had made an early attempt to address equity by including indices of inequality in the social welfare function, and for society's aversion to inequalities (Wagstaff 1991). But he did not pursue the necessary experimental methods for determining the parameters of the model. These concerns have continued into the 21<sup>st</sup> century. Nord (2001) addressed the criticism that QALYs favour the able-bodied over the disabled by proposing that all life years saved should be equated as one (for life extending interventions). Williams responded to this by suggesting that this approach would mean that interventions that provide fewer overall benefits would be equated with those that provide more benefits (Williams 2001).



Johannesson (2001) has proposed that a solution would be to give the same relative change in QALYs the same weight irrespective of the number of expected QALYs (and controlling for age and gender). The average expected number of QALYs for a patient group would be the average expected for that population of same age/gender, divided by the number of expected QALYs for the patient group before the intervention. The problem is to know, *ex ante*, the number of expected QALYs for any age/ gender group, but this is a technical not a theoretical problem. The idea is consistent with people having equal chance of access to resources regardless of current health status, to resources going where ability to benefit is greatest, and does not discriminate by age or gender.

The fact that there is no agreement on how to tackle equity in the measurement of QALYs and that the analysis is not undertaken in a standard way leads to inconsistency in reporting of the findings, as demonstrated in recent reviews (Sassi, Archard et al 2001, Sculpher 2002, Black and Mooney 2002). The consideration of ethics and equity in QALY analysis address important issues for policy but create new problems by moving the debate away from its theoretical origins in preference satisfaction.

These arguments are extra-welfarist, as the utility-based approach would allow for different individuals' preferences for interventions or health states or for differential values to benefits accruing to disabled and able-bodied people. Welfarist approaches can incorporate almost any arguments into the social welfare function to reflect equity considerations, for example 'process utility' (McGuire, Henderson and Mooney 1988). This approach is not as restricted as the extra welfarist methods where only health outcomes can be considered, but it is also not as practical.

The problems that extra welfarism can solve by focussing on units of health rather than having to measure preferences create new problems that do not arise in the welfarist approach which can take account of unobservable arguments in the utility function (Dolan 2000). The extra-welfarists might get around the problem of interpersonal comparisons of utility (by assuming that all QALYs are the same for all people) but are restricted to the consideration of health dimensions only. No account can be taken of people's differences in preferences and willingness to trade health for other forms of welfare.

### **Health -related versus global measures of quality of life**

The debates about how to generate QALYs have to some extent avoided the issue of how the outcomes of health care intervention are defined and determined. Supporters of the CBA have argued that it provides the broadest approach to defining utility that should be the

default approach adopted in decision-making (Mishan 1998). Any approach that is narrower than this has to set out the limitations of what it can measure as an outcome.

It is important to make a distinction between health-related quality of life and quality of life more generally. Quality of life can be defined in multiple varying ways, from vague definitions of "whatever the individual defines it to be," "ability to lead a normal life" and "self-actualisation," to more thoughtful and considered definitions such as those that emphasize social, economic or personal aspects of life (Zhan 1992, Bowling 1997). The World Health organisation has defined quality of life, as distinct from health-related quality of life, in the following manner:

*"[quality of life is] the individuals' perceptions of their position in life, in the context of the cultural and value systems in which they live and in relation to their goals, expectations, standards and concerns."* (WHOQOL 1993)

A more existential definition of quality of life has been given as:

*"the extent to which an individual is able to achieve security, self-esteem and the opportunity to use intellectual and physical capabilities in pursuit of personal goals."* (Engquist 1979).

The concept of quality of life is difficult in scientific research since it is by its nature a relative concept (Aksoy 2000). Some commentators have also considered the ethical and political implications of using quality of life as an outcome measure, especially in areas of health care that cannot be easily defined., for example, the amount quality of life might an individual might sacrifice for society as a whole (Dean 1990). Some of the problems in defining a broad concept of quality of life may be reduced if the concept of health-related quality of life only is the outcome of focus, although this can lead to an excessively narrow view of what quality of life might be (Harris 1988).

The broadening of the concept of the value of health care to include social and psychological aspects of well-being has made more progress in some disciplines than in others (Bowling 1995). Health status, on the other hand, distinguishes between health-specific determinants of quality of life from other determinants (income, job security, living conditions, and the personal and wider social worlds). Health-related quality of life (HRQoL) is usually defined as functional capacity or physiologic functioning. There has been some confusion between this concept, wider concepts of quality of life, and economic ideas of utility in the quality of life literature (Smith, Avis et al 1999) . It has been argued that this confusion began primarily when funding agencies began to require quality of life measures. The broad meaning of quality of life has tended to be brought closely in line with the clinical view, neglecting ethical

and societal dimensions (Killian and Angermeyer 1995). Researchers took health status, disease symptom, and functional questionnaires, renamed them quality of life or health-related quality of life and administered them because they were already available (Smith, Avis et al *ibid.*).

Preference-based (utility-based) measures of quality of life used in economic evaluation have focused on describing and valuing a range of health outcomes or states. QALYs are in essence health-related quality of life outcome units and do not incorporate the wider definitions of quality of life described above (Brazier et al 1999, Blumenschein and Johannesson 1996). There are programmes of research underway to develop approaches to QALYs that measure concepts of social well-being using choice-based approaches (Netten, Smith et al 2002). But in general, quality of life as defined here has not been incorporated into QALY analysis. Instead, a broader evaluation framework using monetary valuation approaches has developed.

### **Re-emergence of cost-benefit analysis and contingent valuation**

There are positive and negative aspects to all current approaches to economic evaluation. The reconsideration of CBA techniques that can incorporate wider considerations of welfare beyond the health dimension has arisen as a response to the continuing problems of defining and measuring quality adjusted life years, and as a way of returning to the theoretical foundations of welfare economics (Olsen 1997).

The problems inherent in the process/ quality of life measures adopted by CEA and CUA may be no less onerous than those in CBA. CBA also has some advantages over the other methods. All costs and benefits are translated into the same units and are therefore comparable (unlike CEA). The value of outcomes to individuals is reflected in the monetary valuation and reflects social value of interventions (social value is the sum of individuals' value) (O'Brien 1996, Klose 1999). CBA acknowledges that health is only one argument in an individual's and (society's) utility function. Interventions can be compared across health interventions but also with other uses of public resources, and in theory with private consumption. CBA incorporates all values that are deemed to be important by society, not only the values that are known and counted *ex ante*. This means that more complex interactions between individuals' preferences, although not known, can be incorporated in the analysis (Morrison and Gyldmark 1992).

Continuing problems with measuring comparative outcomes in CEA and CUA methods have led to a renewed interest in CBA as a methodology in health economics (Johannesson and

Jonsson 1991, Hutton 1992, Zarnke et al 1997, McIntosh et al 1999, Donaldson et al 2002). The attraction of CBA is its theoretical grounding and the return to the roots of welfare economics by including preferences and values in analysis (Boardman 2001). One health economist called this a return to a “mode of thinking “ in economic terms (Gafni 1997b).

Since there are no prices for many health care products or interventions, it is necessary to find alternative ways of undertaking CBA studies. Methods have been developed to estimate value by asking individuals' willingness-to-pay for health interventions rather than by observing people's demand for services and different prices (Donaldson 1990; Diener et al. 1998; Klose 1999). Respondents are asked to provide answers to hypothetical health scenarios about how much they think they would be willing to pay (or willing to risk, or willing to accept in compensation) to move in and out of health states. These are therefore stated preference, rather than revealed demand methods of monetary valuation.

All the methods of obtaining estimates of monetary valuations where there is no market are controversial. In the 1970s, this had already been recognised by those proposing the CBA approach (Mishan 1988). More recent commentators have questioned the basic idea that preferences, as expressed through some monetary valuation, are the same as values (Shiell et al 1997). Equally the idea has been challenged that individual's preferences for health care are stable over time (Shiell and Seymour 2002) and not constructed by the experience of health care (Slovic 1995). Nevertheless approaches have been proposed and employed in empirical studies. These ideas will be examined again in chapter 4 in the context of axioms of choice experiments. The next section reviews these methods and the problems that have been highlighted in the literature.

### **Contingent valuation**

There are a number of approaches to assessing respondents' valuation of different outcomes, and these methods are known collectively as contingent valuation. Contingent valuation (CV) is a survey-based approach for eliciting the public's valuation for health care for use in CBA (Jones-Lee and Hammerton 1985). It involves respondents evaluating, in monetary terms, goods or services that may not be directly measurable. It estimates respondents' stated preferences rather than their preferences as revealed through their consumption of commodities (Morrison and Gyldmark 1992).

The willingness-to-pay (WTP) approach is grounded in Paretian welfare economics and the axioms of consumer sovereignty. This means that WTP can be used as a measure of the marginal benefit of an intervention, leading to a social welfare function which is the sum of

individuals' explicit and implicit willingness-to-pay. The methodology is rooted in the principles of Pareto optimality and the Hicks-Kaldor compensation principle. Whether contingent valuation studies violate the impossibility of interpersonal comparisons of utility is a debated point as raised earlier (Blaug 1997, Williams 2000). There is some strong dissent in the health economics literature about the idea of benefits of health care being translated into money (Carr-Hill 1991, Burrows and Brown 1992). This will be discussed further on.

### **Willingness-to-pay**

The most common approach of contingent valuation studies in health care is the 'willingness-to-pay' that asks individuals to state the maximum they would be prepared to pay to be indifferent between a stated outcome and their current circumstance. They are asked the amount they would pay to reach a specific improvement in health or the amount they would need to be compensated to move to a worse health state (a willingness to accept study). The method incorporates some notion of sacrifice such that an intervention is only of value if a person is willing to give something else up (measured in money) in order to have it. Methods of eliciting these values can be obtained by asking directly (within a range or open-ended); by standard gamble techniques, visual analogues, time-trade-off and more recently, stated preference choice experiments (also known as discrete choice experiments or conjoint analysis).

The WTP approach has been considered to be a more theoretically correct measure of outcome with respect to consistency with welfare economics (Morrison and Gyldmark 1992) and its validity and reliability have been investigated and found to be acceptable as a measure of health state preference (O'Brien 1994). Proponents of this approach have argued for its superiority over other outcome methods. This is due to its adherence to welfare economic principles (Gafni 1998, Birch et al 1999), its strength in considering aspects of care which are traditionally difficult to identify and measure (Gibb et al 1998) and as a way of involving consumers more closely in decision-making for resource allocation (Ryan and Farrar 1995).

One of the first applications of the WTP approach was to assess preferences for NHS hospitals and nursing homes. The results suggested that the group that preferred nursing homes could compensate those preferring hospitals and still be better off, suggesting that nursing homes were the more efficient alternative (Donaldson 1990).

The majority of applications of CV have been used in 'simple' interventions, for example in drug therapy studies (Blumenschein and Johannesson 1999, Johannesson et al 1991) and

dentistry (Matthews et al. 1999). A recent review of CV techniques found 48 studies in health economic evaluation using these methods (Diener et al 1998). The majority of the studies (42/48) valued benefits in terms of willingness-to-pay (revealed demand), with the remainder as price/ demand studies (observed demand). A review published three years later identified 78 studies using WTP. The review considered the legitimacy of the main arguments for WTP as a “superior” tool for economic evaluation. These were given as: the adherence of WTP to welfare economic principles (Johannesson 1996), the inclusion of arguments beyond health in the utility function rather than health only (e.g. Donaldson and Shackley 1997) and the use of commensurate units for costs and benefits necessary for assessing allocative efficiency (e.g. Drummond et al 1997). The authors of the review dismissed the first argument against WTP (insufficient adherence to welfare economic principles) is an insufficient condition for superiority since this would depend also on whether the values implicit in the evaluation approach were the same as those of the society where the evaluation takes place.

The second two arguments were assessed in light of the published evidence. Only 17 of the studies described health states in more than one health dimension. Twelve studies included a comprehensive description of the scenario to be valued prior to the exercise with face-to-face interviews. The authors suggest that only for these studies could valuation of health care beyond a primary health outcome be extracted. Only 25 studies compared costs and benefits directly. However, the fact the many of these studies were only partial valuations meant that it would not be possible to reach a conclusive result directly from these data.

Studies to explore the construct validity and test-retest reliability of CV studies established that WTP is highly correlated with standard gamble methods (seen as the gold standard) for eliciting health state preferences (O'Brien and Viramontes 1994). Two further studies (Kartmann, Stalhammer et al 1996, Farrar and Ryan 1999) have established that the order in which questions are asked does not appear to affect the final results, suggesting that the WTP values are robust within the context of the experiment. A study of the relative sensitivity of WTP and time-trade-off techniques to changes in health status found that WTP was more sensitive to differences in quality of life between different levels of health (Smith 2001).

The issue of establishing the stability of preferences and face validity of these techniques has been a barrier to their implementation (Johannesson and Jonsson 1991, Burrows and Brown 1992). The data are inconclusive as to whether willingness-to-pay questionnaires correspond with respondents' actual willingness-to-pay if there were a real market for these goods and services. A study on the willingness-to-pay for a pharmacist-led asthma

programme suggested that hypothetical willingness-to-pay (under survey conditions) overestimated revealed willingness to pay (measured by the payment to enter the programme) although their findings were not conclusive. A later study, examining literature in areas where there was a possibility for comparison between stated preference and revealed preference showed that WTP was significantly correlated with real WTP (Liljas and Blumenschein 2000). A more recent Nigerian study has also demonstrated that the proportion of people who said they were willing to pay for insecticide treated bed-nets corresponded to people's actual purchase of the bed nets when they became available (Onwujekwe et al 2001).

Outside the health economics literature, a recent study of the valuation of goods to reduce the risks of environmentally induced lung cancer has been able to assess WTP values against actual purchasing behaviour by households to reduce the risks (Kennedy 2002). Although the sample was not the same, it came from the same area with the similar mean income, health states and age/sex profile. The study found that the WTP values and the revealed preference values did in fact converge. These results differed from previous evidence that has found that market values (defined as travel costs for example) did not converge with WTP estimates and travel cost estimates were 25% lower than WTP estimates (Clarke 2002). Some of the reasons given for the divergence of the values in this study were the inclusion of use values and non-use values such as altruism in the WTP estimates.

### **Criticism of the willingness-to-pay approach**

There has been some considerable criticism of CV techniques in health care (Burrows and Brown 1992, Bala et al 1998) both in the methods and in the theoretical foundations of the approach. Morrison and Gyldemark (1992) found that the empirical evidence did not show superiority or inferiority of WTP over quality of life-based approaches to measuring outcomes for economic evaluation of health care interventions. At that time, they judged that there was not sufficient empirical evidence to address whether the advantages of the CBA approach could be realised by this methodology.

One study compared methods for eliciting preferences (including WTP) for health related outcomes for shingles and found no significant correlation of methods across individuals (Bala et al 1998). A more recent exercise was undertaken to compare three different methods of eliciting direct willingness-to-pay (Donaldson 2001). Willingness-to-pay for an individual's own care did not discriminate well between respondents previous ranking of alternatives. It was reasoned that respondents were comparing their care with doing nothing rather than an alternative intervention. If there were a real difference in utility between

interventions, this method would not have identified it. The willingness-to-pay by an individual for two different interventions showed that respondents were guessing the price of the intervention rather than estimating its worth to them (giving higher values to the intervention they believed would have higher prices). The third method, asking respondents their willingness-to-pay to receive their preferred intervention (in their ranking of alternatives) rather than a less preferred option, performed the best of all three methods. The author pointed out the coincidence that the marginal approach is most strongly linked to the concept of a Pareto Improvement (the Hicks-Kaldor criterion that a change should be made if the benefits to some outweigh the costs to others). The study also examined whether ability to pay had an important effect on willingness-to-pay and whether adjustments based on weights attached to WTP values of respondents with different incomes had an effect on the final result of the study. They found that, in their small sample of 79 individuals, these effects were not important.

This study illustrates important aspects of the willingness-to-pay approach. First, it suggests that the general population may find these kinds of questions difficult to respond to and may consider price rather than value. Second, that the marginal utility of money (and therefore differences in willingness-to-pay) among different income groups may be a problem for this form of valuation. Third, that WTP values are not direct utility values (even if theoretically they are meant to be). Like QALY estimation, WTP is a proxy for utility, rather than a direct estimation of it.

Gafni (1998) reviewed many of the objections surrounding willingness-to-pay methods and has categorised them into those that question the feasibility of the approach, the theoretical underpinnings, those that misunderstand the purpose of the method, and emotional arguments. Gafni argues that much of this argument distracts attention away from the proper scientific debates about the importance of the theoretical foundations for economic evaluation. Other commentators have also argued that emotional arguments have confused the debate about the appropriateness of the WTP method (Birch et al 1999). However, it has been argued that these emotional arguments are not as important as the problem of individuals considering the price of health care (Ryan 1999) and their valuations being influenced by their ability to pay rather than their need (or preference) for the intervention (Donaldson 1999). There is some evidence that preferences differ systematically between income groups in WTP studies (Donaldson 2001).

A UK Treasury report into the valuation of public services highlighted the disadvantages of seeking monetary valuations of health (Cave, Cunningham et al 1993). The first is described



by the authors as “policy or strategic bias”, which arises when respondents believe the information they provide could influence policy, and therefore they tailor their responses to achieve a desired policy outcome. The second has been called “property rights bias” (Mitchell and Carson 1989) but is described in the Treasury report as “politicisation bias”. This refers to the influence of respondents’ views about the appropriateness of having to pay to acquire a service. One way around this might be to ask respondents their willingness to accept compensation for not having a change in policy. A review in the Treasury report of WTP and WTA studies in the public sector found WTA values tend to be significantly higher than WTP. This finding was shown in a review of controlled experiments in environmental economics using both methods on the same respondent group (Cummings et al 1986). The Treasury report dealt with this issue in detail, exploring whether there is some systematic bias that affects WTP and WTA unequally, or whether individuals do not value goods “correctly”, by maximising their utility. It concluded that the WTP method should be preferred since the likelihood of *status quo* bias and the additional incentive for “protest responses” is higher for WTA. The relationship between WTP and revealed preference (where markets and hence prices exist) was also explored in the Treasury report, that concluded that the evidence was inconsistent and therefore: “it would be unwise for the time being to place much reliance on the monetary valuation aspects of public services.” (ibid. pp 71) This point is raised more generally by Willan and colleagues (Willan, O'Brien et al 2001) whose theoretical and empirical evidence suggested that WTA is about twice as much as WTP in health care.

Elsewhere, arguments against the direct elicitation methods have been even more critical of these methods of contingent valuation. In the environment sector in the US, contingent valuation methods (both using WTP and other methods of valuation of outcomes) have been proposed as a means of determining the size of financial compensations that should be paid out by companies. This compensation might either be to the government or to individuals for the damage caused to natural resources by corporations (Arrow, Solow et al 1993). An article published by the editors of the Harvard Law Review (1992) strongly criticised this method of valuation as biased, unreliable and speculative, so much so that that the costs of using these methods outweighed any of the benefits (in terms of their use in legal compensation proceedings). The fundamental problem outlined in this article was that hypothetical answers to WTP questions had no cost associated with being wrong, as they would in real life. Therefore, there would be no incentive for respondents to make the mental effort to be accurate. The article also pointed to the problem of strategic bias where respondents may purposefully misrepresent their valuation, and the problem of values being

constructed for the first time during the survey leading to arbitrariness of valuation. What this amounts to is a violation of the axioms of rationality in decision-making.

Another key argument made in this paper was that estimates for vastly different quantities of resources tend to fall within a similar range: the WTP to preserve one acre of wilderness was similar to that for 100 acres. The authors proposed that WTP was more commensurate with an “imaginary gift to charity” than a reflection of a market value of a non-market commodity. “People decide whether a cause is worthy, then pick a nice number to donate to that cause” Harvard Law Review Editorial, pp 1993)

This position is countered (also in the field of environmental economics) by the arguments that CV methodology is more sophisticated than simply asking individuals their direct WTP for an environmental resources and that bias in survey design is a problem for all surveys and not CV alone. Methods of data collection are more likely to ask the public “If it cost \$X would you pay for it/ vote for it?” (Hanemann 1994). What people value should be left up to them and reliable and replicable methods of assessing these values is still a worthwhile activity.

These debates remain unresolved and confusing. It is hard to establish whether the difference of viewpoint as to the acceptability of WTP and contingent valuation methods concerns the feasibility and applicability of these methods of evaluation in the real world (in which case they can be improved), or whether the arguments are based on the theoretical validity (and should be abandoned).

### **Considering the value of care “beyond health outcomes”**

Over the past ten years, the argument has been put forward that practitioners, researchers and policy-makers need to consider the value of health and social care beyond the narrow confines of clinical effectiveness. In one of the first articles to consider this in specific relation to health, Mooney and Lange suggested that evaluation (and policy decisions) have often failed to take into account the full impact of particular health care interventions that may have a wider impact on utility than narrow definitions of health gain (Mooney and Lange 1993). In a later article, Mooney developed these arguments by suggesting that the value of health care is more than the sum of its contribution to health; that there is value in knowing that others are cared for, and in knowing that these services will be available (and not subject to ability to pay at point of need) when they will be needed in the future (Mooney 1994). This debate is pertinent to the thesis since Mooney has argued for a fuller investigation of what patients want from health and social care services (Mooney 1998). Furthermore the idea that

people as citizens may have interests in preserving certain characteristics of the health care system even if these may conflict with the cost-effectiveness evidence is highly relevant to the debate on how services such as PDC should be valued. Mooney's contribution is also to reflect on the predominant interest of researchers in outcomes as a technical problem of identification and measurement. It has also been suggested that outcomes are not technical entities but incorporate value that are unavoidable since the decision to focus on specific health outcomes and not the wider contribution of health care is itself a value judgement (Shiell, Hawe et al 1997).

Ratcliffe and Buxton have explored the idea that there are values other than narrow measurements of health gain. They published a study on liver transplantation using choice modelling techniques that suggested that patients were prepared to make trade-offs between the chance of a successful transplant and improvements in the quality of care they experienced (Ratcliffe and Buxton 1999). For services that fall outside the mainstream health care system this evidence is informative, especially where it is highly unlikely that a full recovery is achievable (and is not expected by the individual). There is a strong argument that it is not only health gain that is valued by patients attending the service. Furthermore, these arguments suggest that individuals might trade-off some aspect of individual health gain health against other aspects of health care provision such as equity of access or other aspects of the quality of health care. This may be particularly relevant to services such as palliative day care where health gain may be only one of many important aspects of the experience. This is one of the key questions explored in the empirical analysis in this thesis.

There has been progress made in developing wider definitions of quality of life for older people (the main recipients of palliative care services). There is currently a research programme underway funded by the Economic and Social Research Council that focuses on the quality of life for older people (Bowling et al 2003). In palliative care, there have been developments in the outcomes literature that attempts to incorporate concepts such as spirituality (Brady, Peterman et al 1999, Efficace and Marrone 2002) and hope (Herth 2001) in studies of the benefits of care. There has also been critical debate about the concept of quality of life that has shifted the interest from symptoms to functioning. This incorporates adaptation and relativism in quality of life research that is seen as more relevant for people with long-term or palliative care needs (O'Boyle 1997). These developments are extending the meaning of quality of life beyond narrow definitions of health, but they are mainly taking place outside the health economics literature. The debate within the health economics literature has been concerned more with the examination of the values that are placed on

health states (who should value them and how) rather than with the meaning of the health states themselves.

### **The emergence of 'communitarian' arguments in the health economics literature**

Gavin Mooney and colleagues have argued that there ought to be a re-examination of the basis for valuing the benefits of health care and allocating health care resources (Mooney 1998). Mooney was drawing on ideas developed outside economics about the "claims" that communities have on resources and on decision-making that are different from those made by individuals maximising preferences (Mooney 1998). He argues for a re-examination of the basis for valuing the benefits of health care and allocating health care resources. Earlier writings on communitarianism emerged in the disciplines of philosophy and social science as a critical response to the publication of the book by John Rawls in the early 1970s which argued that the principal task of government is to secure and distribute fairly the liberties and economic resources individuals need to lead freely chosen lives (Rawls 1971). The central arguments of what became 'communitarianism' (to contrast it with Rawl's view of liberalism) distinguished between three sorts of claims that a community might have over that of individuals decision-making: methodological claims about the importance of tradition and social context for moral and political reasoning, ontological or metaphysical claims about the social nature of the self, and normative claims about the value of community (Avineri and de-Shalit 1992). Etzioni has been a key proponent of these ideas and has written about them in the context of politics (Etzioni 1993).

In health economics, communitarianism has helped to shape what is seen as an alternative way of conceptualising consumer involvement in decision-making. This is seen as an alternative to basing evaluation on the consumption preferences of individuals, which assumes a narrow, consequentialist notion of benefits. Culyer suggested that evaluation and policy decisions have often failed to take into account the full impacts of interventions, in particular the benefit of knowing that others are cared for and for knowing that decision-making is equitable (Culyer 1980).

Mooney has argued that communities (defined geographically, socially, and within institutions) ought to be involved in and expressing preferences for the overall framework within which individual decisions about health care can be made. Without the specific knowledge of the consequences of health care interventions, individuals may not be capable of expressing fully formed preferences. Instead, communities ought to have a role in defining the values and norms for decision-making within which technical or operational decision about health care can be made.

The operationalisation of communitarianism within a health economic evaluation paradigm is still being developed and tested in empirically based studies (Jan 1998, Mooney et al 1999, Jan et al 2000). Shackley and Ryan (1995) have suggested that community or consumer involvement in decision-making should be separated into two distinct activities: the level of deciding whether a specific service should exist, and, once it has been decided that such a service should exist, what it should consist of. These debates are very pertinent to the themes of this thesis that explores ways of valuing complex services once it has been decided “by society” that such services should be supported.

One of the problems with moving away from what is straightforward to quantify and measure is that the ideas of ‘community’ are themselves complex concepts. The definition of the term ‘community’ (who its members are, how they should express their views, how decisions ought to be taken) has not been overtly specified in the literature because it can take many forms. However, it could be argued that the basic concept of community is a contested notion in political and economic thought. The idea that there is such a concept as a community of common interest (rather than divisions of class, gender, race, social status and other forms of identity) may be more clearly definable in some settings than in others. Mooney has argued, quoting from Sen that it is “better to be vaguely right than precisely wrong” (Mooney 2001, pp42). However it is not clear that the communitarian argument is any more “vaguely right” than other sources of value (such as individual preferences for the outcomes of health care), since there is no clear gold standard against which to measure these alternatives. Communitarianism appears more intuitively close to a democratic approach to evaluating health care. But problems of defining the “community” and of defining and agreeing the decision rules when different communities do not have the same values does not avoid the problems that standard economic evaluation using the CEA/CUA approach faces. It has been argued in this chapter that the CBA approach can, in theory incorporate all the known and unknown arguments in an individual’s utility function. This could include preferences for “communitarian” or democratic processes and constitutions as well as preferences for the health of others as well as oneself. The criticism of welfarism is that there is a tendency to conceptualise individuals’ preferences as preferences for an individual’s private benefit (in terms of health, money, or any other contributor to welfare). However, this argument is not fully supported in the theoretical arguments of Mishan and others set out earlier in this chapter.

### **Application to complex services – which methods could be used?**

To decide which approach would be the most suitable for evaluating any health care programme, it is clear that the simplest method that can meet the objective of the study

should be adopted. This follows the rule set down by Occams' Razor that one should not increase, beyond what is necessary, the number of entities required to explain anything. It admonishes the researcher to choose from a set of otherwise equivalent models of a given phenomenon the simplest one. Furthermore resources are wasted using more complex research methods that will not provide better information for decision-making.

Also, it is necessary to be clear about theoretical foundations of the research and whether strict adherence to theoretical principles will add to the validity of the findings. Since all methods of evaluation make value assumptions and all use proxies for outcomes, all methods are a trade-off between the data that it is possible to collect and validity of the findings (in terms of adherence to welfare economic principles and ability to interpret the results).

In some settings, this matters more than in others since the relationship between an outcome measured in natural units (e.g. the hip replacement example described earlier) and the value of the intervention were it to be purchased in a perfect market, are commensurable. For health care settings that are more complex, there are reasons for hypothesising that the relationship is also more complex: the measurement of observable changes in quality of life and the value of those changes to the individual experiencing them is not clear and may be highly individual to each person. Non-acute health and social care interventions could fall into this category.

In health and social care settings where the benefits are not well defined, the subtle nature of the benefits may not be discernable using quality of life instruments that only identify larger changes in quality of life. It can be argued that the QALY approach was not designed to measure overall welfare and does not discriminate well between different forms of psychosocial health, which is the most obvious domain where complex services might make a difference in quality of life. The fact that the EQ-5D instrument considers anxiety and depression in the same domain (i.e. as synonyms of each other) illustrates this point since, in palliative care for example, research has indicated that these psychological states are distinct and should be addressed in different ways (Craven 2000, Hinshaw 2002, Ly et al 2002). However this has not been tested empirically and it might be useful to do so in a palliative care context to test whether these arguments are valid.

For interventions that address wider dimensions of welfare, a form of cost-benefit analysis might be an appropriate way of considering the value of a complex service, but this method too has problems. The first is the issue of having to separate the measurement of

willingness-to-pay from ability-to-pay. In the context of many health and social care services, this is a major methodological problem. The concept of 'willingness-to-pay' goes against this philosophy of care in many settings where services are provided free of charge.

Another issue is the possibility that people who use these services may be either too frail or too vulnerable to take part in a survey that requires complex decision-making and judgement. This may be an important concern in research in these areas. Alternatively, there is a chance that respondents might 'game the system' by offering responses that are different from their genuine valuations in the belief that this will have an affect on the care (or the cost of their care). The use of a direct monetary valuation technique could meet significant "politicisation bias." There might be lack of will to support for this type of research from the providers of care and prevent it from gathering any empirical data at all.

In the context of palliative day care these are all important concerns. Socio-demographic data from the only comparative study of PDC in the UK (reported in the next chapter) showed that the mean age was 74 years in this group (range 50-74), with only 4.5% of attenders in paid work. More important, only 20% of carers had any paid employment (full or part-time) and 63% of patients did not live in privately owned accommodation. There is a danger, therefore, that patients would say they would choose not to attend palliative day care at all if they asked what financial contribution they would be prepared to make. This raises the issue of whether a service has 'value' if users are not willing or able to make any sacrifice in order to attend palliative day care. The nature of the question may need to be refined to identify some other form of sacrifice other than financial to assess the consumption foregone that respondents would be prepared to sacrifice in order to attend PDC.

However, there are innovative methods being developed that may provide a way forward in this direction. Choice experiments, where patients trade between different attributes of care (of which one may be price or a proxy for price but it does not have to be) may provide a possible way of adopting a quasi CBA approach. This is an indirect technique of contingent valuation. The approach has other important properties that might also be relevant to the evaluation of the types of services considered here since it explores the value of specific attributes of care rather than the service as a whole. In order to examine whether this would be an appropriate method and whether a QALY type approach could be undertaken, a review of the theory of choice experiments, methods and the issues they raise is the subject of chapter 4.

## SUMMARY

This section has reviewed economic theory relevant to the arguments in the thesis. It showed that economic evaluation of health care has been dominated by methods that have moved away from the cost-benefit analysis approach, which is the closest to the practical application of welfare economic theory. This has been due to the great difficulties in valuing the outcomes of health care in monetary terms and also stems from a rejection of the idea of using individuals' preferences for deciding the value of public services. This is perceived as an inappropriate and inequitable criterion for making decisions about the distribution of public resources.

The QALY approach makes assumptions about the nature of time as linear and additively separable which is more problematic in some contexts than in others. In the context of palliative care interventions, time may be inherently different at the end of life and valued differently, from time earlier in life. It also incorporates a narrower definition of quality of life than may be relevant and meaningful in particular care settings. The separation of health benefit and the valuation of that health benefit would seem to be an important barrier to using the health gain method for interventions that occur later in life. This argument that objectively measured health gain can be different from the subjectively felt value of health gain may also hold for people with long-term illness who do not expect to reach full health. Therefore, there are persuasive arguments for a preference-based approach to evaluating complex services.

The next chapter aims to put these questions more firmly in the context of current and past evaluative studies undertaken work in palliative care, since this is the example on which this thesis is focused. This bases the discussion clearly in the real world of evaluation, and the actual problems faced in undertaking this type of work in complex areas of health care. The review assesses what is known about palliative day care from the economic evaluation literature and other evaluation sources, and to highlight the gaps in the literature. The previous palliative day care study is presented as well as evidence from other areas of health and social care. The previous palliative day care study and evidence from other areas of complex health and social care has informed the decision about how to make progress in the empirical work presented in this thesis.

Other types of evidence are also explored in order to conceptualise the nature of the outcomes of palliative day care, and to consider what type of evaluation would be appropriate to make progress in the evaluation of palliative day care. This work informs the empirical work presented later in the thesis.



## Chapter 3

## **Review of the economic literature in palliative care research**

### **Introduction**

The purpose of this review is to assess how far the issues raised in the previous chapter on the theory of economic evaluation have filtered through to the economic evaluation of palliative care. The argument made here is that economic evaluation of complex health and social care services is less advanced than the evaluation of other health care intervention because of the problems of defining and measuring outcomes. These problems are not specific to palliative care but are relevant to other types of health and social care where outcomes cannot be easily defined by one simple measure of outcome to produce a cost-effectiveness ratio. Evidence from qualitative literature on palliative day care is also presented since the empirical study presented later in the thesis draws on this work.

The second half of this chapter addresses the problem of measuring outcomes for interventions that are complex and multidimensional and where outcomes are hard to conceptualise and measure. Examples from other health and social care sectors where similar issues of measuring outcomes might arise are discussed. Lessons are drawn from the first economic evaluation of palliative day care, and insights from studies in other areas of health and social care are presented.

### **SECTION 1. REVIEW OF THE ECONOMIC EVIDENCE IN PALLIATIVE CARE STUDIES**

This review is presented to demonstrate that the empirical work presented in this thesis represents a step forward in testing methods of economic evaluation in palliative day care research. Since the body of evidence on the cost-effectiveness of palliative day care is very limited, the review of the evaluative literature considers evidence from a wider range of palliative care studies. It demonstrates that the economic evaluation studies in palliative care have not in the main been well done and do not answer important questions for economic evaluation in this area. An analysis of the methodologies used in these studies demonstrates an absence of economic theory underlying these studies.

This section critically reviews 14 studies of palliative care services that have included economic analysis. The inclusion criterion was studies that considered resource use or cost as well as the outcomes of palliative care. The key themes explored and critiqued in this

literature review were: how outcomes have been conceptualised and measured; how outcomes have been considered alongside costs; the robustness of the findings; and the interpretation of the findings by the authors.

### **Data sources and review criteria**

The review of economic studies was embedded within a systematic review undertaken for the Welsh Office. The search was undertaken by trained systematic and the economic papers identified from that review were passed to HRD for data extraction and review. The economic review formed one chapter of a report to the Welsh Office completed in 2001. The review was updated by HRD for the thesis using the HEED and NHS EED databases up to August 2002, but no new papers were identified.

The systematic review was undertaken using the databases MEDLINE, CINAHL, CancerLit, PsychInfo, EMBASE, PallCare Index, EPOC register, System for Information of Grey Literature (SIGEL), Applied Social Science Index (ASSIA) and Sciences Citation Index (SSI). Studies were identified from the data-base inception to end of 1999 and updated using Medline, CINAHL and PsychInfo to end 2000. The following key words were used: palliative or hospice, terminal care, terminally ill, palliat\*, hospice\*, dying\*, end of life; and effective or evaluate, random, methods, economics, statistics, trends, organization, utilization; and service or team. This was augmented by hand searching Palliative Medicine, Progress in Palliative Care, the Journal of Palliative Care and the Journal of Pain and Symptom Management to end 2000, examining references from papers retrieved and a search of the grey literature.

For inclusion, studies must have compared palliative care or hospice teams with conventional care (present or historical). An intervention was defined as two or more health care workers, where at least one had specialist training or worked principally in palliative or hospice care. Study populations were patients with a progressive life threatening illness and their caregivers (defined as family, friends or significant others). Usual care was routine community and general hospital/oncology services. Outcomes were classified as: pain and symptom control, quality of life and death; patient and family satisfaction / morbidity pre- and post-bereavement. Non-English language articles were translated. Anecdotal and case reports or studies without measured outcomes were excluded.

### **Overview of the studies**

Appendix A shows a summary of the evidence from the papers in the systematic review. The search criteria did not include specific treatments or specific groups of patients, but was

defined as any service carried out by a multidisciplinary palliative care team with at least one individual having specialist palliative care training, or who spends all their working time in palliative care, working in a hospital, home/community setting or in a combination of these. Articles were excluded if they reported individual case histories, or if they described any interventions not usually considered to be part of palliative care, that is, surgery, chemotherapy, radiotherapy, or anaesthetic procedure. Articles were also excluded if they reported a review of palliative care without empirical data; a needs assessment exercise with no intervention described or evaluated; a description of a palliative care intervention but no evaluation; a qualitative study with no comparative element; developing countries interventions (where the context of care is markedly different from the UK), or an intervention deemed not to be palliative care, for example disability, rehabilitation or chronic pain.

In all, 43 studies were found that contained potentially relevant economic data. Studies that contained primary economic analysis represented 14 out of the 43 studies. All the papers that contained some comparative evidence of the costs and resource consequences of palliative care and measurement of outcome were considered for inclusion. Of these studies in the review, nine of them evaluated a home care intervention (Vinciguerra et al 1986, Cummings and Hughes 1990, Tramarin et al 1992, Hughes 1992, Bloom 1980, Zimmer et al 1984, McCusker et al 1987, Ventafridda et al 1989, McCorkle et al 1989). This included services provided by multidisciplinary teams, hospital-based home care, physician-led home care, nurse-led home care, and "home hospice". The others were inpatient hospice service, either alongside home care (Kane et al 1984, Greer 1986, Dunt et al 1989) or as a stand-alone service (Axelsson and Christensen 1998). The only other type of service evaluated was a nurse coordination service (Raftery et al 1996). None of the studies considered palliative day care services.

The research methodologies can be categorised as follows: those which presented economic analysis alongside a randomised controlled trial (five studies); those which presented health care utilisation data *after* having failed to establish significant differences in clinical outcomes from an RCT (two studies); economic data gathered alongside an observational study (five studies); and analysis of health care utilisation and monetary value of health care only (two studies).

### **The conceptualisation and measurement of outcomes**

In 12 of the 14 studies that considered outcomes alongside costs, outcomes were measured using health-related quality of life (HRQoL) instruments to measure health gain. They did not include measures that reflected preferences for health interventions or health states. The

types of instruments that were used focused on specific aspects of ill health associated with patients in a palliative phase of illness (anxiety, depression, social dependency); on the health of older people (who make up the majority of the palliative care population); or general health profile (the Sickness Impact Profile, used in Zimmer 1984). All the studies except three (McCusker 1987, Bloom 1980, Axelsson 1998) used at least three different outcome instruments. The three papers that did not incorporate quality of life considered no outcome measures at all and evaluated programme costs only.

All the studies were undertaken before a specific palliative care outcome instrument had been developed and tested. The palliative care outcome scale (POS) was validated and published in the late 1990s (Hearn and Higginson 1999). All the studies reviewed here were undertaken before this time. The problem the researchers faced was to incorporate a range of clinical and health-related quality of life outcomes into the evaluation in a way that would be sensitive enough to detect important differences.

The scope of the studies has been narrowly defined. The majority of these studies are small, single institution studies. The resource use data is focused on the providers of care and how much might be saved by changes in service re-organisation. No comparative measure of outcome is reported. For example, Vinciguerra and colleagues (1986) undertook a resource use analysis of their pilot oncology home care programme in Italy, describing the types of resources that were used for this programme (e.g. clinical equipment, staff involved) and presented a comparative financial analysis and per diem cost savings of this programme. The per diem savings reported were very large (34%). These savings were reported from the programme and did not consider the impact of the costs of care on patients' use of other resources or on the costs to the family of home care. The only outcome considered in this study was financial cost to the provider that meant it was not a full evaluation.

The focus on provider costs was also adopted in two other studies (Axelsson 1998 and Bloom 1980). One of these (Bloom 1980) did include some qualitative evidence of outcomes from carers that suggested that patients had a better quality care at home, but this was not developed further in this paper. All the single institution studies suggested that palliative care was less costly than conventional care even though cost data were not rigorously collected and the numbers of subjects small, that is, fewer than 200 subjects (Axelsson 1998, Raftery 1996, Vinciguerra 1986, Tramarin 1992, Ventafridda 1989, Dunt 1989, Bloom 1980). These studies are similar to phase one clinical trials. They are all small trials or observation studies that aim to understand whether there is any clinical benefit at all resulting from a new

intervention. They do not consider whether the size and scale of this benefit is economically worthwhile.

Only one study (Tramarin 1986) attempted to construct a single quality-of-life instrument. This was a study of a home care intervention for people with AIDS. It was a very small prospective study of 10 subjects, randomly selected from 17 eligible patients. The authors used a Quality of Well-being Scale (QWB) that expressed outcomes in terms of complete well years adjusted for diminished quality of life. The QWB scale incorporates physical functioning and symptoms and an overall quality of well-being score graded from 0 (death) to 1 (asymptomatic optimal functioning). The authors modified this score by recalculating the time frame from "well-years" to "well-weeks" and produced an overall ratio of cost per well week. No other information was given about this score, its validity in this population cohort, or how a "cost per well week" ought to be interpreted. No details were given either about how the overall weighted score was derived, and the patient scores were not published. The authors describe this as a cost-utility analysis, but it is clear from their evidence, that this was not a utility-based score but a method of adjusting (in an obtuse way) for reduced health along a few domains that were not clearly described. This is an example of a small-scale study that attempted to consider a combined measure of length of life and quality of life but the methodology was not well demonstrated or clearly based on cost-utility analysis theory (either the health gain or preference-based approach).

All four randomised controlled trials focused on a single palliative care unit (Zimmer 1985, Kane 1984, Hughes 1992, Cummings 1990). These studies had more rigorous study methodologies but costs and outcomes were not considered together in a comprehensive economic evaluation. The focus of the economic analysis was still provider costs with outcomes analysed separately. The studies found no difference in health-related quality of life between palliative care and non-palliative care patients except that there was increased satisfaction with place of care among palliative care patients, an outcome found in all the studies. This is a relevant finding as it implies that satisfaction with care, which is fundamentally about patients' self-expressed preference for where and how they are cared for, may be a very important outcome of palliative care.

Analysis of the discussion at the end of these papers is revealing: the attention is on clinical outcomes and costs, and the issue of increased satisfaction is not addressed, certainly not in relation to amalgamating patient preferences into an economic evaluation. One reading of this could be that increased satisfaction (or realisation of preferences) is not a 'hard' enough outcome in effectiveness research. Whether or how to incorporate this phenomenon into an

economic evaluation, and how preferences and values might be incorporated into economic evaluation, has not been addressed in any of the studies published in this area.

### **The economic focus on cost**

The over-riding concern with the costs of care is shown by the aims of these studies. Cost-effectiveness is conceptualised as demonstrating lower costs of care for palliative care services. Authors of the papers have highlighted the need to demonstrate cost savings from palliative care as the main purpose of their work. This reflects concurrent concerns in the USA especially about whether medical insurance should cover palliative care services (Kane 1984, Cummings 1990, Greer 1986, Hughes 1992). It also reflects concerns about perceived inadequacies of local provision of care for people who were dying with the focus on a new service (Zimmer 1984, McCusker 1987, Dunt 1989, Raftery 1996, Tramarin 1992), the need to provide data to purchasers on the cost of a new service (Axelsson 1998, McCorkle 1989); and the need to demonstrate that palliative care does not equate to higher costs (Vinciguerra 1986, Bloom 1980, Ventafridda 1989).

Costing a new service for the first time is a worthwhile activity. However these studies do not address questions of whether an intervention is economically important, that is whether additional costs or savings of providing palliative care are associated with additional or reduced quality of care. This evidence therefore represents a narrow approach to economic evaluation. None of the studies conceptualise economic benefit as a way of expressing outcome. Only two studies demonstrated superiority of the intervention, one in HRQoL scores (Ventafridda 1989) and one in pain score (Dunt 1989). Three studies demonstrated that palliative care led to higher patient satisfaction with care (Vinciguerra 1986, Hughes 1992, Zimmer 1984). The first two were randomised controlled trials of home care versus hospital care. The latter study considered additional physician-led support versus usual home care.

### **The consideration of outcomes alongside costs**

The previous section showed that all the studies in this review of palliative care that have included outcomes alongside costs have used evaluation instruments that reflect a range of quality of life dimensions. This method of evaluation has been called a form of 'modified cost-effectiveness analyses' (Goddard 1989). Goddard suggests that this approach to measuring quality of life in the palliative care setting reflects the problem of measuring health-related quality of life for individuals in the final stages of life. The limitation of this for economic evaluation is that without a single unit of outcome there can be no incremental cost-effectiveness ratio (Drummond et al 1997). This presents a problem in assessing the

overall effectiveness of different alternatives and identifying a decision rule for assessing the superiority of one intervention over the other. Either one of the interventions in the comparison must be more effective across all domains of quality of life (and across all HRQoL instruments), or across only one domain where all other domains are the same. If the findings show that one intervention dominates another across some dimensions of outcome but is dominated in others, then no decision rule can be applied since it is not clear whether one intervention is better than another overall.

Chapter 2 outlined the particular welfare economic rules on which economic evaluation is based: the decision rule based on Pareto optimality is that the benefits to the gainers of any change in policy have to be greater than the losses to the losers. If the gainers can potentially compensate the losers and still be better off, then a change in policy is a Pareto improvement. This is the decision rule on which economic evaluation is constructed.

Considering how decision rules have been made in these studies is enlightening. One approach that has been taken in one group of these studies (Ventafridda 1986, Cummings and Hughes 1990, Raftery 1996, McCusker 1987, Axelsson 1998) has been to argue that they have established clinical equipoise between palliative care and conventional treatment and then reported reduced costs afterwards. These can be categorised therefore as cost minimisation studies. The palliative care intervention is dominant and therefore difficult questions about the marginal costs and benefits do not arise. A Pareto improvement has taken place since there are only gainers and no losers. However since these studies established no evidence of difference rather than evidence of no difference, the argument that there is true clinical equipoise is not strictly true. The research instruments and sample size used to establish differences in outcomes may not have been powerful enough to have measured real differences, or the result may have been obtained by chance.

**Box 3.1: Study designs of economic studies included in the review, with overall findings. 'Equivalen' (E), 'More satisfaction' (S), 'less pain', 'None' refers to the overall finding for the outcomes of the study**

<b>RCT</b>		<b>Resource use only reported from an RCT</b>	
Kane (1984)	E	Cummings (1990)	E
Hughes (1992)	S	Raftery (1996)	E
Zimmer (1984) *	S		
McCorkle (1989)	E		
<b>Observational Studies</b>		<b>Costing study in framework of an observational study</b>	
Vinciguerra (1986)	S	Bloom (1980)	None
Ventafridda (1989)	S	McCusker et al (1987)	None
Greer (1986)	E	Axelsson (1998)	None
Tramarin (1992)	E		
Dunt ( 1989)	Less Pain		
<b>All studies</b>			
Population average: 391 subjects			
Range: 42 – 1874 subjects			
Median: 167 subjects			
Only one multicentre trial included in this review			

*\*two papers from the same study by Zimmer et al et al*

### The validity of the findings

Criteria are now well established for publication of economic studies (Weinstein et al 1996, Drummond et al 1997). However these had not been widely disseminated when these studies were published since many were undertaken in the 1980s. Furthermore only one study (as far as could be determined from authors' title and address for correspondence) was undertaken by a health economist. To some extent this may have had some influence over the types of studies that were undertaken and how the economic evaluation was designed.

One of the main forms of potential bias in small, non-randomised studies is clearly selection bias. The concern about the observational studies is that patients who are enrolled in a palliative care programme have different preferences from patients who have conventional care (Tierney 1994, McWhinney 1994). These preferences could be for less aggressive treatments, to avoid inpatient care and to be independent for as long as possible before death (Greer 1986). There is some evidence for this in the literature. Patients in palliative care settings tend to have been ill for longer, have better social support network, are more likely to be white, and to be younger than patients who do not enrol in palliative care programmes (Emmanuel 1996). The randomised controlled trials included in this review support that when these factors are controlled for by randomisation, differences between intervention groups disappear.



In all of the papers, possible sources of bias are discussed. Problems that are reported are not related to the economic data but are limitations of the overall study design. These were non-representatives of the study sites (Cummings 1990, Raftery 1996), non-representativeness of patient groups (Cummings 1990, Kane 1984, Bloom 1980, McCusker 1987, Axelsson 1998), group contamination (Kane 1984), differences in case-mix between control and intervention groups (Cummings 1990, Raftery 1996), and small sample sizes (Kane 1984, Bloom 1980). Problems are reported in these economic studies that are not exclusive to economic evaluation but are limitations of the overall study design. The problem of the lack of data on costs to the family was discussed by the authors of four studies (Ventafridda 1986, Tramarin 1992, Bloom 1980, Axelsson 1998).

Given that studies are generally undertaken by enthusiasts keen to demonstrate cost savings, it is important that five studies either failed to demonstrate anticipated savings or cautioned readers in interpreting their results. Cost studies have usually been undertaken in order to demonstrate that a new programme is either cost-neutral or cost saving. It may be argued that those authors undertaking a cost study alongside an efficacy trial already anticipate that a new programme will reduce costs of care. This could be described as a form of researcher bias. Null findings suggest that the cost savings are more tenuous than the authors have concluded.

While some papers reported very large differences in resource use between control and intervention groups, these tended to be the weaker papers in terms of study design (observational studies rather than randomised controlled trials) and sample size. Only two studies suggested that the reader should interpret the results with caution. However, the majority of authors have gone to some length to try and explain the differences in resource use between intervention and control groups, and to assess the possible sources of bias introduced into the study.

### **Discussion of quantitative studies**

This review of empirical economic research found that none of the 'economic' studies published in this field had incorporated values and preferences for different forms of palliative care. Their focus has been narrower and has evaluated the impact of palliative care on the overall costs of care. The problem with a cost- focussed approach in the wider context is that it does not provide any evidence for whether the additional costs or savings of a change in policy represent a good use of resources. It is interesting that research into a service where the user is at the very centre of the decision-making process (in deciding whether to accept home care, agree to be admitted to hospice, accept conventional care

only) has not considered the problem of including patients' preferences in economic evaluation up until now. This may be because researchers have not had the knowledge, understanding or expertise to apply more robust economic measures of outcome in this area.

One of the problems of this narrow approach was identified even before the majority of these empirical studies were undertaken:

*"The real problem is to determine when and in what ways a consideration of costs is reasonable in administrative decision-making [...] What the administrator may view as a pattern of wasteful expenses for statistically minuscule benefits the clinician whose practices are being examined may consider a pattern of justified expenses for a series of individual treatments, each undertaken in the best interests of the particular patient"* (Bayer 1983)

These days, the 'clinician' might be substituted for the patient but it is argued that the idea (and the problem) is the same.

### **Practical and ethical problems in quantitative evaluation of palliative care**

Previous trials of palliative care services have described both practical and ethical problems and a high refusal or uptake among potential patients (Rinck et al 1997, Grande et al 2000a, 2000b). In a systematic review of 11 randomised-controlled trials of palliative care, two studies problems were so severe that no results were reported (Rinck et al 1997). In 10 studies there were problems with patient recruitment. There were also problems reported problems of population homogeneity, patient attrition, defining and maintaining contrast between interventions, and selection of outcome.

Zimmer and colleagues (who undertook one of only two RCTs of palliative care) have argued that their study has demonstrated that an RCT is feasible for a multi-dimensional intervention (Zimmer et al 1985). But that these services have to be added to or change to what is seen as normal care. Goddard (1993) has also argued that the RCT is the preferred design for evaluating alternatives in health care as it can eliminate various forms of bias. However, there are various ethical problems with randomising patients who are dying. This is made easier if patients are randomised to two forms of care, rather than service/ no service.

Goddard put forward the case that a randomised trial in the context of palliative care services would ensure that differential outcomes were due to different models of care rather

than differences in patients but recognised that RCTs would be difficult to organise (Goddard 1989). The difficulties relate to the nature of the illness and the needs of the patients to be cared for now rather than at some time in the future. Randomisation to no service would be perceived as unethical. For example Higginson (1999) found that GPs were reluctant to agree for patients to be randomised to the 'no support' arm in a trial of palliative home care. It is easier if two different models of palliative care are evaluated.

Back in the 1980s, Ward (1984) undertook a study of nurse-led palliative home care, and reported that the reliance on nurses who may be "protective" towards patients in a palliative phase of illness slowed down arrangements for undertaking a detailed cost and outcome study for a sub-sample of patients. In this period, a number of patients died or were admitted as inpatients before any data could be collected. The small number of patients eventually recruited meant that no comparison of outcome could be made with conventional care. Ward also reports problems of completeness of diary data by relatives of patients who were dying which meant that detailed comparisons of costs were not assessed. Jarvis and colleagues in their later study of a specialist palliative care programme incorporating home care, inpatient care and support services, also found that a large number of patients had to be excluded from the study due to the acute nature of their illness (Jarvis et al 1996). Also, the authors report that the proportion of patients dying before data from two time periods could be collected (60%) meant that the sub-sample who survived were not representative of the populations.

Jarvis and colleagues (1996) have argued that an important source of bias in palliative care studies was that some patients (often the most ill) do not participate, and this may be a large proportion of people who could benefit from a palliative care programme. In their study, they found that only 34/84 subjects had data for two time points for comparison. The reason was a marked physical deterioration in the respondents. This raises the question of the internal validity of these studies. Undertaking the valuation at an earlier stage in the disease trajectory could improve the completeness and reporting of results but this would not capture changes in health-related quality of life of the acutely ill patients.

This is not only a phenomenon in palliative care research, but has also been found in a more general review of home based support for older people (Elkan et al 2001). This review found an absence of evidence of improved health and functioning states that was explained by the authors of the review as due to the patients with the poorest health dying and therefore dropping out of the analysis. This meant that outcomes could only be measured on the

subset of patients who survived longer. The research tools were not sensitive enough to detect modest improvements in health or functional ability.

## **SECTION 2. ECONOMIC EVIDENCE IN PALLIATIVE DAY CARE RESEARCH**

Since palliative day care is a recent development in palliative care, research evidence of its effectiveness and cost-effectiveness is still in its very early stages. Studies that have been published have been small descriptive studies of individual services and surveys. There has also been a recent collection of essays on palliative day care that brings together experience and practice of delivering palliative day care services (Hearn and Myers 2001). Although this is not peer-reviewed evidence and does not directly compare palliative day care with other forms of care, this descriptive analysis is invaluable for understanding the process of palliative day care, and the context in which evaluative studies are undertaken.

A review of the evidence for palliative day care was undertaken in the late 1990s (Spencer and Daniels 1998). The review traces the development of day hospice in the UK using predominantly descriptive evidence and analysis of recent UK policy. The review presented various aspects of palliative day care (such as service provision, access and meeting consumers needs) and provided a broad picture of the kinds of provision that is available in the UK and for whom. However, at the time of publication there was almost no evaluative literature. The review describes some of the attempts to try and quantify or describe the process of palliative day care and the extent to which meets its objectives. None of the papers cited considered cost-effectiveness directly but were attempting to conceptualise how palliative day care might be evaluated given the heterogeneity of the intervention.

Limitations in terms of identifying appropriate outcome measures were considered to be an important reason why evaluation would be difficult to achieve. The purpose of the first review of studies by Spencer and Daniels was to set out a framework for considering how palliative day care might be evaluated. It did not consider in any depth how patient outcomes might be affected by the range or intensity of day care services, or the levers by which palliative day care might increase or decrease the cost-effectiveness of care, nor how patient preferences could be taken into consideration in any evaluative framework.

### **Palliative day care studies – qualitative evidence relevant to the economic evaluation**

Published evidence of the effectiveness of PDC is very sparse. A few descriptive studies have been undertaken to either describe a single service in the UK (Hopkinson and Hallet 2001, Faulkner, Higginson et al 1993), or in the USA (Thompson 1990); to describe the ethos and principles of day care (Corr and Corr 1992, Kennett 2000); to set policy and

service standards for palliative day care in the USA (Olson 1989) and to report stakeholder views on implementing a new palliative care service (Lohfield 2000). These studies have been published, in the main, without reference to one another. Nevertheless, they reflect a large degree of similarity in defining the need for PDC and in describing the structure of the service as it is currently configured or could be organised in the future.

Only one descriptive study, by the author of this thesis, was undertaken specifically to inform an economic evaluation. This was a pilot study undertaken in five PDC centres in Greater London to assess the structure, process and outcomes of palliative care. This was a participative observation study where the researcher spent a week in each centre. Additional observation data, financial data and interviews with stakeholders informed this study (Douglas et al 2000). Some of the characteristics that key differences in provision and philosophy between PDC centres are presented below:

The qualitative evidence has informed the design of the economic evaluation of PDC as it provides insights into the general problems of evaluation in this context. A number of themes have been drawn out from these descriptive studies. These relate to the heterogeneity of intervention, the heterogeneity of patient needs, the complexity of describing PDC as a single intervention, how to relate inputs to outcomes, and finally whether user and provider perspectives show a different viewpoint in terms of what is important about PDC. These themes are explored in this chapter.

**Table 3.1 Characteristics of the five PDC centres in the North Thames study**

Centre	A	B	C	D	E
Attendance per week	24	45	49	36	70
Philosophy	More medical	More medical	More social	Medical/ social	More social
% NHS funded	98%	27%	41%	22%	24%
Managed by	NHS	Voluntary sector	Voluntary sector	Voluntary sector	Voluntary sector
Location	Inner city	Inner city	Suburban	Suburban	Suburban
Inpatient unit	attached	attached	Not attached	attached	attached
Day for younger people	No	Yes	Yes	Yes	Yes
Discharge policy	Yes	Yes	Yes	Yes	Yes
No. therapists (paid)	11 (10)	11 (3)	10 (6)	11 (3)	8 (6)
Transport provided	Taxi service	Volunteers	Volunteers	Volunteers	Volunteers
Medical staff	Appointment	Appointment	Appointment	On-call, emergency only	On-call, emergency only

### **Heterogeneity of intervention**

The qualitative studies demonstrate that the breadth of possible activities is diverse. Thompson (1992) describes a very wide range of services offered on one American day hospice. These are described as water-based activities for symptom control (whirlpools, swimming pools, and baths), through to personal care (bathing, hairdressing, and beauty), hot meals, spiritual activities, arts and creative activities, and help with activities of daily living (walking, dressing, cooking, and cleaning). Faulkner et al (1993) found that across 12 PDC centres in one UK region, specialist services were provided by doctors, nurses, aromatherapists, hairdressers, beauticians, physiotherapists, art therapists, social workers, occupational therapists, the chaplaincy, reflexology and massage therapists.

The only example of published guidelines for PDC has been written in the 1980s in the USA (Olson 1989). The authors of the guidance set down standards that palliative day care providers have to meet in order to claim reimbursement. They were to make provision of the following services: medical care, nursing care, social work, transportation, nutrition, and alternative educational and supportive therapies such as music therapy, art therapy, stress management, family counselling, grief counselling, guided imagery training, recreation therapy, speech therapy, occupational therapy, physical therapy, nutritional counselling, and relaxation training.. This is a long list of activities, the efficacy of which has not been fully established using rigorous trial methods. In the UK, there has been no policy of common standards established. This has meant that PDC has been provided in different types of accommodation and with different levels of staffing and other resources. (Faulkner et al 1993)

Each of the qualitative studies has argued that a well-designed PDC service should be able to offer patients a range of services, incorporating flexibility into the programme. The emphasis in both North America and the UK has been on allowing patients to decide autonomously what their needs are and on trying different activities.

Faulkner (1993) also reported that different UK centres emphasised different aspects of PDC. One centre did not provide any medical, nursing or physical input. One centre did not provide any creative activities, each believing that it was not the role of PDC to provide these services that could be accessed elsewhere or were not part of the therapeutic role of PDC. Douglas et al (2000) found that there was a difference in philosophy or emphasis in day care provision across centres, with some emphasising the clinical surveillance role more than the creative opportunities provided by PDC. In the UK, there is no statutory guidance on what should be provided, or on the effectiveness of individual therapies. Each centre is guided by

local circumstances and the ideologies of the providers of the services whether these are the services that the patients need or want has yet to be established.

### **Heterogeneity of patient needs**

Two studies have focused on the ability of PDC to meet different patients' needs with different kinds of services. Thompson et al showed that one PDC centre provided different kinds of care on different days of the week to accommodate sub-groups of patients (Thompson 1990). For example there was a day for attenders who do not feel too ill and still need the care offered by the day centre. For this group, an artist was present and the centre was more oriented around creative activities. Other days were quieter with less input for those patients who are either more ill or more elderly. A separate day is also provided for younger patients. This pattern was also observed in the UK studies.

Douglas and colleagues found that what was offered in PDC varied from one centre to the next. For example, the emphasis could be on creative activities, psychosocial support, individual or group activities, rehabilitation, medical treatment or nursing care (Douglas et al 2000). The emphasis was on patients having some control over what happens to them, and on deciding week by week, what activities would be beneficial. Older patients appeared to want fewer active therapeutic services; younger patients (under 65 years) said they wanted more control over what they received and they were willing to change their package of care more often. Younger people who attended appeared not to stay as long in the day as older people, and said they were more likely to attend the centre for a particular treatment or consultation, rather than to use the social facilities of the centre.

In all centres, patients were referred for a range of reasons: for a specific palliative care need such as continued symptom management or nursing care, psychological support or for dealing with complex family situations; for short-term breaks for carers and for social and creative opportunities for patients and their families. In some centres, patients undergoing radiotherapy or chemotherapy were referred for additional support. Patients usually attended once a week, although this depended on the patients' needs and use of other services.

This meant that no patient would necessarily receive the same package of palliative day care services. The intensity, timing and range of services and therefore the possible impact of these services on quality of life will differ between patients, or patient groups. The ability to benefit from palliative day care and the extent to which patients were actively involved in other forms of care and social activities also differed between patient groups, and between people at different stages of their illness.

### Complexity of describing in the intervention

The authors of qualitative and descriptive studies of PDC have attempted to describe the service in a straightforward way that is amenable to a wide readership. PDC has been described as *“both a health experience and in many ways even more importantly – a psychosocial experience or personal experience”* (Corr and Corr 1992, pp156). Other authors have described it by its service components in order that providers of the service have clear guidance on how the service should be structured (Olson 1989). Another author summed day care up as *“a day that begins with self care, is then generally spent in work, and ends in relaxation”* (Tigges, in Corr and Corr, 1993: pp 171). This sounds straightforward, but does not explain the purpose of the service and why patients are offered a range of activities.

All the authors of descriptive studies have argued that PDC is more than the sum of its parts. *“The purpose is more important than the activity. Helping individuals retain a sense of integrity and dignity in the face of progressive losses is the primary goal of hospice and hospice day care”* (Thompson 1990, pp 30). This position is supported by another study that suggested that *“elementary craft activities guided by an occupational therapist or simple class exercises led by a physical therapist have an importance that may not be apparent from superficial observation”* (Corr and Corr 1993, pp 157).

Thompson goes on to describe PDC as, *“a centre [that] provides a therapeutic milieu where patients can receive the necessary support, training and equipment to resume activities and adapt to changes in previous life roles, bringing meaning and value to patients’ remaining days”* (Thompson 1990, pp 30). While this is a more elaborate description, the means by which PDC *“brings meaning and value”* to patients is not well described, and the lack of description of day care from the patients’ point of view underscores this.

Authors have described a *“warm and supportive atmosphere”* (Faulkner et al 1993). Douglas (2000) also reported that the extent to which palliative day care differed across centres and from other kinds of health services was a matter of degree. For example, it is a more diverse service than most other statutorily funded health services. While people presented with obvious, visible care needs, they also may attend day care because staff feel they live in situations of acute social isolation and where no other form of care is offered. There are high staff-to-patient ratios in PDC centres and the workforce is often dependent on voluntary as well as paid therapists and support staff. The effect of this on the provision of services is to both make it more flexible and adaptable to individual patients’ needs but also to make it less stable and more reliant on the willingness of local people to support the service. These are



important observations since most forms of evaluation assume that there is such thing as an average intervention and average effect.

### **User perspectives in the description of PDC and measurement of outcomes**

The studies described so far have all been undertaken by providers of PDC (Corr and Corr 1992, Thompson 1990) or by independent observers of the service (Faulkner, Higginson et al 1993, Douglas, Higginson et al 2000). The impact of this service on the users themselves, as perceived by them, has not been a focus of qualitative research until very recently.

A phenomenological study to assess patients' perceptions of PDC found that patients could be divided into one group who could tolerate living with cancer without much change to their daily living, and another that felt that this life event required some considerable adaptation (Hopkinson 2001). The author reported that these different kinds of needs for palliative care (one more intensive, one less so) could be supported in PDC since care focused on individuals' needs and was flexible enough to manage both. What this study did not explore was what *specific* aspects of PDC enhanced or detracted from patients ability to cope with living with cancer. Nor did it explore the feelings of people who felt less satisfied with the care they received (although it was reported that not all interviewees felt as positive as the majority, in a cohort of just 12 patients). The over-riding perception of patients was that it is a "humanistic" service, which is a term that is open to different interpretations and not a straightforward outcome for evaluation.

One study that did consider the effectiveness of a specific aspect of PDC was undertaken in one hospice in South London (Kennett 2000). This study explored the experiences of patients taking part in the creative arts. The author was involved in the provision of this service. Again, no negative aspect of participation in this activity was reported, and themes are all expressed in an extremely positive light such as self-esteem, autonomy, hope, and social integration. This study illustrates an important problem in evaluating PDC: the need for providers to demonstrate effectiveness in the face of financial constraints, and the lack of external evaluation by outside researchers. The one qualitative study (undertaken as part of the North Thames study described in the next section) interviewed PDC attenders over a three-month period (Goodwin et al 2002b). Interviews were undertaken outside the day care setting, usually in patients' own homes, and the researcher did not work in day care. The study reported negative as well as positive statements by patients. The most important aspect of PDC was found to be "meeting other people" and "the company", regardless of the type of centre attended or the range of services offered. Patients who attended the centres with an emphasis on medical input were more likely to say that specific medical care were

important to them. This suggests that patients' preferences are not complete when they arrive but are constructed by their environment. Furthermore, there was some dissonance between the providers' view of PDC and what the patients said they wanted, as evidenced by this quote from a patient:

*"I only attend for the counselling. I've tried relaxation and I can't do that.... I'm not really interested in art and craft... I have to come all day just for the counselling"* (Goodwin et al 2002b, pp 278)

### **Issues in the economic evaluation of complex interventions**

This thesis has set out to investigate the general problem of evaluating the outcomes of interventions that can support decision-making within an economic evaluation framework. The specific problem is how to evaluate outcomes where interventions are complex. A definition of complexity that will be used in this thesis is proposed here. This complexity has been conceptualised here as interventions that cannot be pre-determined. They will depend on the specific needs of the individual whose needs will change over time. These individuals' needs will be different from the needs of others in the same phase of illness, due to their life circumstances and previous experiences. Services are interactive; they depend not only on the context of provision, but also on the person receiving the care. They cannot be assumed to provide the same value or benefit to individuals who are perceived (objectively) to be in the same health state. This means that the relationship between inputs and outcomes is inherently variable. Furthermore, the benefits of PDC are complex as they relate to individuals' ability to meet their goals and expectations and to remain positive in the last months of their life.

The nature of these goals and how PDC meets individuals' needs may not be seen in the same way by providers and the users of these services and their carers. Across different services, the emphasis or philosophy of care differs in subtle ways, revealing different institutional beliefs about the purpose of PDC services, and these are not universally agreed by providers. Other interventions, such as community mental health care also have this characteristic of intrinsic complexity, and are "*purposively flexible*" to match the needs of individuals with the services offered (Byford and Sefton 2002).

The argument made in this thesis is that the nature of complexity is not confined to palliative care. The lack of consensus described by Evers and colleagues also describes the problem faced by researchers in palliative care. Sectors where the needs of users of the service are

diverse and the intervention are not easy to define account for a wide range of services provided within NHS and social service care.

The “complexity” described in the nature of PDC may also be found in other health and social care sectors. Examples of other sectors where similar issues arise are include many forms of nursing, mental health care, long-term care, care of the elderly, and in the organisation of services that cross between health and social care, where interventions are not discrete and are designed to meet a multiple of needs.

**Box 3. 2 A sample of the interventions that could be described as ‘complex’:**

- Organisational interventions to meet long-term mental health needs
- Cross-sector interventions to meet the welfare needs of people who are marginalised by society, such as homeless people, new economic migrants and asylum seekers
- Interventions that have multiple aims, such as strategies to reduce the effects of poverty or to improve educational achievement and attainment
- Interventions that are provided to the families of people with life-limiting illness (for example counselling and support by clinical nurse specialists)
- Interventions where efficacy has not been established through clinical trials, such as counselling, and complementary therapies, and where the evidence is contested by different groups.

A database search to identify other areas of “complexity” is not straightforward since the language to describe this complexity will differ across studies. A search of the HEED database under “complex” found only one study which was a review of the evidence of the cost-effectiveness of intervention to address the needs of people living with diabetes (Gulliford 1997). The review drew similar conclusions that the challenge of defining and measuring the effectiveness of the intervention was a major challenge for economic evaluation, and that the methods of organising care may be the important factor in determining the overall cost-effectiveness of care. Further searches under “organisation” in the same database found 23 studies. The studies defined as having some of the same characteristics as PDC were those of interventions that provided broadly social as well as health related care provided in non-standardised way to people with a range of different needs depending on their life circumstances. Interventions for people with long-term schizophrenia would fall into this category. Economic evaluation evidence in this field has also adopted cost-consequence methodologies (Smylie et al 1991, McCrone et al 1994, Chan et al 2000), focussing on the costs of care and reporting the clinical effectiveness of interventions.

This cost-consequence approach was also followed in economic evaluations of other interventions that focussed on the organisation of care, such as stroke units (Grieve 2000) and asthma management (Sullivan, Weiss et al 2002, Evans, LeBailly et al 1999). The identification of ways of synthesising costs and preference-based measures of benefit into cost-effectiveness ratios has not been a major focus of this body of work. One study has reported the development of an instrument to measure outcomes in terms of utilities for interventions to manage depression (Bennett, Torrence et al 2000), but no published studies were found to date that had employed this instrument in economic evaluation.

In their review of economic evaluation of mental health, Evers and colleagues highlighted the problem of identifying an appropriate unit of effectiveness as a major problem in mental health care studies. This is impounded by what they characterise as a lack of consensus about the aetiology and appropriate treatment for many psychiatric illnesses (Evers et al 1997).

Norton (2000) has described the characteristics of long-term care as differing from health interventions care that focus on treatment and cure-oriented care. He describes four distinct characteristics of long-term care: care is focussed on people with 'chronic' illness with no expectation of full recovery; services are located outside mainstream provision and funding (funded through the private sector or voluntary organisations); informal (unpaid) care plays an important role, and finally (in the USA) private insurance does not cover the costs of care. Norton's analysis does not consider the problem of measuring the cost-effectiveness of these services *per se*, but the analysis does provide insights into why these kinds of services are different and the nature of the complexity facing evaluators.

### **Evaluating palliative day care as a complex intervention**

The argument developed so far in this chapter has suggested that interventions such as PDC that aim to meet the needs of patients with complex problems are inherently difficult to evaluate. This is because of the problems of identifying appropriate units of effectiveness and the lack of focus on preference-based measures of outcome in published studies of these types of interventions. In these sectors, it is also difficult to identify and isolate particular services that contribute uniquely to goals associated with finding meaning and hope and meaning different things to different client groups (Spencer and Daniels 1998).

Therefore it can be argued that the question of whether PDC and other complex services are "effective" needs to address the more precise questions "what aspect of these services are effective and for whom?" If a PDC study found that patients benefited (however defined)

from PDC, it might not provide useful insights into the aspects of PDC that had been beneficial to different kinds of patients. For example, one component of PDC is the team of clinically trained staff attending to patients on a one-to-one basis; another component is the 'safe and friendly atmosphere' that is promoted. If the second characteristic of PDC were more important to patients (and to their overall well-being) than the first, then this would suggest that centres could focus on this (less resource intensive) aspects of care and potentially reach more patients than could be accommodated in a one-to-one service. So the question of how PDC is to be provided, and what aspects of the service are important to patients, is also important for practical and policy-making purposes.

The qualitative evidence is sparse, descriptive and, until recently, was undertaken in order to inform others on the purpose of PDC and how to set up a service. The more rigorous and independent evaluation of PDC inputs and outcomes has only come later. There is an enthusiasm for PDC in the qualitative literature that is enlightening as it demonstrates the level of support that the service has had from the people who provide it, and how this enthusiasm, rather than independent evaluation, may have ensured the continuation of the service.

### **The North Thames palliative day care study**

This study was the first of its kind to undertake a comparative approach to evaluating palliative day care. The study was undertaken in light of the qualitative evidence presented above that indicated that the service would not fit into a neatly prescribed evaluation framework since the intervention could not be described in a way that would lead to straightforward measurement of inputs or outcomes. The economic evaluation was undertaken at the same time as a clinical effectiveness/health-related quality of life evaluation. The results of this study have strongly influenced the empirical and methodological work outlined in the following chapters of this thesis.

Between 1997 and 1999 a multicentre study was undertaken in five palliative day care centres in North and South London. These centres represented different philosophies in palliative day care, emphasising more medical care or more social care, as reported in Douglas 2000, which was a pilot study for the North Thames palliative day care study. The aims of this study were to identify whether one approach to palliative day care was more effective than the other, and whether patients who attended palliative day care had different outcomes over time to patients who did not (who were recruited through palliative home care nursing teams).

Palliative care attenders were recruited at their second visit to day care and interviewed at baseline, 4-6 weeks and 14-16 weeks after first attendance. Data on all inpatient, outpatient, community, and social care activities was also collected at this interview. Two disease-specific HRQoL instruments were adopted and resource use information was collected relating to the last four weeks of care before each interview. Open-ended questions to elicit a response from patients in their own words were also asked at each interview.

A single measure that would represent the outcome of PDC had not yet been developed or validated at the time of the study (and has not to date). For this first evaluative study in PDC, a more diffuse approach had to be adopted. Patients and families' views of, and satisfaction with, the services received were recorded in open comments. Health-related quality of life was measured using established scales validated in palliative care populations. These were a 10-item Palliative Outcome Scale (Hearn and Higginson 1999) and a 16-item McGill Quality of Life Questionnaire (Cohen et al 1995). Both cover overall quality of life, physical symptoms, patient anxiety, fears and well-being. These data were analysed separately from the resource use data.

The main findings of the effectiveness study was that the HRQoL measures could not detect any significant differences in the quality of life between the day care and comparison patients, either in terms of change over time or differences between the groups. No differences were detected between the different models of palliative day care in any items on the Palliative Outcome Scale or the McGill quality of life instruments. The results were similar when patients were divided into different survival cohorts.

The qualitative interviews, however, told a different story. In all centres, most patients reported that they appreciated day care. When asked, "what is day care like" they reported two main positive components, 'getting out' and 'meeting others'. These accounted for over 60% of the reasons for liking day care. At baseline interview over half (54%) of the people felt that they had already changed as a result of day care, and 30% had learnt a new skill. Of these 19% had learnt an art & craft skill, with 5% learning a new exercise and 7% stated that a change in them was another type of skill. At second and third interviews the total number of respondents who reported learning a new skill was 43% and 38% respectively. Most respondents felt that there was no downside to day care (76% of respondents in the first interview and 78% by the last interview) (Goodwin et al 2002b).

The main conclusion was that patients who attended day care expressed satisfaction and said they strongly valued the service when they were interviewed, but that these preferences

did not translate into detectable differences in HRQoL scores. This presents difficulties for economic evaluation. First, the consideration of costs alongside a range of outcomes is an indecisive form of economic evaluation as argued earlier in this chapter. If one outcome measure shows benefits and another shows potential harm of an intervention, then the results cannot be transparently interpreted. Second, since the qualitative evidence reported that palliative day care was beneficial but the quantitative evidence did not, this suggests that the quantitative outcome measures may not be sensitive enough to pick up relevant dimensions of the benefits of day care. There is no unequivocal evidence of the benefit or lack of benefits of the service. In this context, a full economic evaluation cannot be undertaken.

There were also important methodological problems with this study. A randomised controlled trial could not be undertaken as it was seen to be ethically unacceptable to withhold palliative day care from people who could benefit. A waiting list approach (to offer palliative care at a later date) was not seen to be acceptable in a patient group with life-limiting illness if such a list did not already exist in practice. A problem with the comparative study design was the identification of a comparison group. Originally, it was intended to recruit patients through specialist home care nursing services. These patients were intended to be the group of patients who wanted to attend palliative day care but, for particular reasons, were not able to attend a palliative day care centre. The reasons people might not be able to attend were the long distance from their home to the nearest centre, the physical difficulty in leaving the house (living up flights of stairs), as well as those people on a waiting list for a particular palliative day care.

However, as it turned out, there were no patients identified by home nursing teams who would have liked to attend but who were unable to for physical reason (ambulances were able to bring incapacitated patients down stairs), and there were no waiting lists in the centres that took part in the study. This meant that the patients who were recruited into the comparison group were those patients that home care nurses thought would be likely to benefit from PDC but who did not want to attend. There are good reasons for suggesting that this group of patients was different from the group of patients that chose to attend. These differences were not picked up in the demographic data, but there may have been other characteristics of these groups of patients that made them different.

Also, it was decided for ethical reasons that patients could not be interviewed before they attended palliative day care as this might have affected whether they decided to continue. Patients were interviewed within the first few weeks of attending. In that time, the baseline

interview suggested that patients had already improved their outlook. The subsequent interviews may not have picked up additional improvements above those already measured at first interview.

The equivocal results in the outcome study were a major stumbling block in the economic evaluation, leading to a deeper examination of why the quantitative and qualitative results came to different conclusions. This had in part to do with the problems of study design and partly the nature of the intervention and the way patients felt about the changes in quality of life they experienced by attending palliative day care. A critical assessment of the evidence from qualitative studies of palliative day care supported the notion that palliative day care had an impact that was not captured in either disease specific quality of life measures that have focus, in the main, on pain and symptoms or on acute stages of illness.

### **SECTION 3. EVIDENCE OF THE COST-EFFECTIVENESS OF DAY CARE FOR FRAIL ELDERLY PEOPLE**

Since the population who use palliative day care services are usually (though not exclusively) older people, the literature on other day care settings for older people has some relevance. The kinds of challenges faced by researchers in this field and the practical and methodological problems they faced are briefly discussed here as they are relevant to the evaluation of palliative day care. One study has been undertaken on a frail, elderly population using conjoint analysis techniques, but this was not an economic study and was not underscored by a random utility approach (Racher and Kaufert 2000). Another study was also undertaken (again not an economic study) in a day hospital setting (Townend 2000). The economic research evidence is limited in this field.

In a review in the 1980s of the evidence of the cost-effectiveness of day care settings for frail, older people, Gerard concluded that day centres were more cost-effective than day hospitals, given that they provided similar benefits at reduced costs (Gerard 1988). However the data on the specific outcomes of day care was limited in this review, and as such, the relative cost-effectiveness between settings could not be established. In 1999, a systematic review of day hospitals was published, reviewing 12 trials of almost 3000 subjects (Forster, Young et al 1999). The review reported no significant differences in alternative services for elderly patients, in terms of odds of death (preventing death) of patients attending day hospital compared with alternative forms of care (what was described as comprehensive care or home care). Patients in both groups had a similar chance of experiencing deteriorating function (on a range of scores). Eleven trials provided information on costs/resource use. Average inpatient use was reported for all trials. Eight trials compared



treatment costs, but these were not comprehensive (i.e. excluded home nursing costs) for six studies. The evidence suggests that day hospital is either slightly more expensive or similar to alternative forms of care.

The actual findings of the review are of limited use in the palliative day care context as the patient intervention was different. Patients received multidisciplinary assessment and rehabilitation – presumably with a view to some recovery of health status, rather than in the last few weeks or months of life. The main evidence that can be gained from the day hospital review is that it faces similar evaluation challenges. This is because of the nature of the intervention (multidisciplinary, multifaceted, and providing medical and social care, as an intermediary between inpatient and home care) and because of the patients who attend.

The authors note that the control groups receive some other sort of active treatment (comprehensive care). This is similar for palliative day care patients who receive specialist home nursing as well as attending palliative day care once a week. Both interventions (day hospital and palliative day care) are relatively weak interventions in that differences in quality of life between intervention and controls might be hard to detect among the effect 'noise' created by other (potentially more intense) health and social care services.

Second, the day hospital review found that no studies reported a summary statistic to represent a patients' overall health status. Different measures of health status were used and analysed in different ways without arriving at a single overriding message to conclude the review. This is an important aspect of the study: like palliative care studies, the focus has been on the clinical aspects of care, rather than preferences of patients and carers or the burden/ release of resources due to the presence of a day care service. This means that making the judgement about whether these services are worthwhile has been seen as less important than measuring the disease-specific (and study-specific) outcomes needed to make clinical decisions about patient's improvement/deterioration. Similar to the review of quantitative studies on palliative care earlier in this chapter, these studies are asking questions about efficacy and provide little insight into either the effectiveness or cost-effectiveness of the service they focus on, despite the possibility of analysing cost data in the majority of these studies.

The most recent study of adult day care from the USA was a randomised controlled trial of 108 elderly patients to receive day care and 104 patients who acted as controls by remaining on a waiting list for three months (Baumgarten 2002). This study gathered data on respondents' subjective assessment of day care (which was positive among both clients and

their carers), but found that standard research instruments revealed no superiority of day care attendance, and no difference in the cost of care. The authors conclude (in the same vein as other studies reported here) that it is difficult to demonstrate objectively or quantitatively the benefits of day care that have been strongly perceived subjectively by the participants. They make a strong conclusion, given the evidence, that high levels of participation should be incorporated into future programme goals, without the empirical data with which to support this recommendation.

### **OPUS: a measure of social care outcome for older people**

The development of a method of identifying the outcomes of care for older people has recently been disseminated (in 2002) as part of the Department of Health's Outcomes of Social Care for Adults Initiative, and developed by the Personal and Social Service Research Unit (Netten et al 2002). The aim of this project was to design an outcome measurement instrument that reflected the relative value that older people put on welfare resulting from social care services, and to incorporate domains that are of concern to the providers of these services as well (ibid.). The approach to identifying preferences was through choice experiments. The research team undertook extensive pre-piloting and developed a choice experiment to establish the relative importance of the attributes included in the outcome instrument. A sub-set of respondents were interviewed using the same methods but with additional information relating to monetary benefits included in the experiment. From this process, an outcome instrument was developed that included five domains: personal care, nutrition, safety, social participation and involvement and control over daily life. The instrument asked about current levels of unmet need, whether informal carers were involved in meeting their needs, and what level of need they would have in the absence of all intervention. The experiment found that willingness-to-accept estimates were relatively high (around £1300) which they interpreted as surplus benefit over cost for social care. This is an interesting finding in the context of the palliative day care study. Chapter 4 discusses the issue of including willingness-to-pay/willingness-to-accept attribute in a choice experiment, and whether the interpretation of surplus monetary benefit is justified. It considers this evidence in the light of the decision not to include this as an attribute in the PDC choice experiment study.

### **Relevance of the research evidence to economic evaluation in palliative day care**

The review of the economic studies in palliative care has demonstrated that the assessments of the outcomes of palliative care have been narrowly defined with clinical endpoints, rather than economic endpoints. Economic endpoints have only focused on process (e.g. hospitalisation) and not outcome. These studies may answer important

questions specific to the local environment about the likely consequence of a new palliative care/ hospice service on the management of budgets, but they are not very useful in the wider setting, as costs are not considered alongside the health impact of an intervention. Even if these studies had been better designed, none has come close to producing outcomes that could be translated into a more universal measure such as QALYs. It appears therefore that this has not been on the agenda for palliative care research until now.

The second drawback of the cost-effectiveness analysis (CEA) approach in palliative day care is that it cannot incorporate the views of the users of the service. These studies do not provide any evidence for whether changes in a disease specific measure is important, whether shorter lengths of stay in hospital is to be preferred to longer lengths of stay (by patients or their carers). One of the key features of palliative care is that the patient is seen to be at the centre of the decision-making process. CEA studies do not incorporate this into the research design. Lessons from qualitative research evidence suggest that this approach will not suffice in palliative day care research.

The qualitative evidence has provided some tentative reasons why people choose to attend palliative day care and these are reported to be not primarily for clinical care, or for reasons that are definable as "health related". This presents an interesting dilemma. If palliative day care is not producing any identifiable health gains, then should it be provided at all? Taking a societal perspective and leaving aside the question of who should fund the service, the question is whether services should be offered that produce no identifiable 'health gain.' This is not a technical question but relates to the value that people (society) places on health services. If society has a preference for providing care and support to people who are facing death, then the question is not if this should be provided, but how it ought to be provided.

The services provided in palliative day care are not aimed at addressing patients' objective health needs only, but are designed to meet a range of subjectively health and social care problems as defined by individuals. Wider quality of life assessment needs to be adopted that can produce a single index of quality of life, incorporating some notion of preferences for the psycho-social well-being as well as health related quality of life of individuals who attend the service. This would suggest that a cost-utility approach or contingent valuation approach ought to be considered. Also, evidence from day hospitals for the elderly suggests that evaluation will be complex where services are not aimed at addressing patients' objective health needs only, but are designed to meet a range of subjective health and social care problems as defined by individuals.

## SUMMARY

The first section of this chapter reviewed the economic studies on the delivery and organisation of palliative care (as opposed to specific clinical interventions). It considered the nature of the economic evidence that has been published and the strengths and weaknesses of this evidence. It demonstrated that the evaluative studies in palliative care research that have considered economic issues have not incorporated wider consideration of outcomes other than clinical effectiveness, measured using a range a HRQoL instruments. For this reason, they do not capture important dimensions of outcome in palliative care. They do not reflect individuals' values and preferences for particular types of care when they are in advanced stages of illness and possibly facing the end of their lives.

The review considered these papers in light of one of the main arguments of this thesis: that the incorporation of values and preferences in the measurement of outcomes is central to economic evaluation in the context of palliative care, and for decision-making. Economic research in other areas of health and social care faces similar challenges. Examples of other studies that have considered these issues were also discussed.

The final section summarised the arguments and demonstrates the gap in the palliative day care literature and a need for economic studies that can incorporate complex concepts of inputs and outcome, values and preferences. The review has helped to identify why palliative day care was different from many other health interventions and why straightforward economic evaluation would be difficult to undertake in this context. The next chapter discusses the approaches to empirical investigation of palliative day care in light of the literature, and in light of the theoretical arguments set out in chapter 2. It provides the rationale for the methods that were adopted and reviews the strengths and weaknesses of different evaluation methodologies for measuring outcomes in this context.

## Chapter 4

## **Empirical investigation of the outcomes of complex services: theoretical and methodological issues**

### **Introduction**

The purpose of this chapter is to consider how to conceptualise the outcomes of care in complex services and from this to consider the advantages and disadvantages of approaches to economic evaluation in the specific context of complex services. It considers different approaches to undertaking evaluation using well-known simple approaches, and more complex methods. The issues are applied to palliative day care (PDC) in light of the evidence presented in the previous chapter.

One of the key themes of this thesis is the distinction between the health gain and health preference approach to evaluating outcomes of health and social care. Both approaches are explored here. Chapter 2 reviewed the theoretical foundations of these methods and concluded that the closest method to the application of welfare economics was contingent valuation, a preference-based approach. However, deriving a monetary valuation for the benefits of health care presents important methodological and practical difficulties. Chapter 3 reviewed the literature to date that has attempted to evaluate palliative care and palliative day care (PDC) in particular. The review found that the only comparative study undertaken in PDC did not show any significant differences in outcome between those attending PDC and those who did not using palliative care (disease-specific) health gain instruments, although this study was methodologically flawed. It also demonstrated that none of the published evaluation studies adopted a utility-base or a contingent valuation approach to measuring the outcomes of PDC.

### **SECTION 1. SUMMARY OF HOW THE EVALUATION OF COMPLEX SERVICES IS DIFFERENT FROM OTHER HEALTH SERVICES**

The evidence presented in chapter 3 suggests that complexity in health and social care is related to the heterogeneous characteristic of these services since people attend for different reasons. They may face different life circumstances and consequently may need a different range and intensity of specialist care and general support. Qualitative evidence suggests that the heterogeneity of the intervention and heterogeneity of users' needs means that it has more in common with social day care for older people, and mental health day care than

with acute health care interventions that address the physical health of individuals (Wimo and Wallin 1990, Victor and Higginson 1994, Marshall and Crowther 2001). The themes highlighted in box 4.1 below summarise the reasons why the evaluation of complex services such as PDC is different from the evaluation of mainstream health services. This reasoning has led to the consideration of different ways of undertaking economic evaluation methods. These provide different kinds of evidence of the effectiveness. This reflects the contrast between these services and others that are well-defined, one-dimensional interventions that are amenable to clinical trial type evaluation (Douglas and Normand 2001).

**Box 4.1 Challenges to applying health care evaluation techniques to complex services based on the research evidence presented in chapter 3.**

Health services are relatively straightforward to evaluate if they have the following characteristics:

- They are highly structured, specific interventions, delivered in mostly the same way for all patients. *(Complex services cover social, psychological, spiritual and existential aspects of quality of life as well as physical health);*
- They focus on improving physical health, with specific care pathways or algorithms to follow. *(Complex services do not aim to provide the same pattern of care for all patients. The focus is on the patient as an individual);*
- The outcomes are clearly defined in terms of improvements in physical health and length of life, and there are uncontested, accepted definitions of positive and adverse outcomes. *(Positive and adverse outcomes are not well defined and may even be contested between the different professional groups. Complex services can be relatively less intensive interventions: for example patients may only access a service once a week alongside other clinical or social services);*
- The focus is on the patient only and usually on a specific aspect of their health *(Complex services focus on wider social or pastoral needs and the well-being of the whole family).*
- *There may be specific difficulties in economic evaluation for interventions that occur at the end of life. These relate to the short amount of time a person has left to live and the (usually) worsening health state they experience, relative to an expectation of a full healthy life. Since any improvements are relatively small and over weeks or months rather than years, a quality adjusted life year saved may not be a sensitive or meaningful unit of outcome (this is explored further in the text in this chapter).*

## **SECTION 2. OVERVIEW OF THE CONCEPTUALISATION OF OUTCOMES FOR COMPLEX SERVICES**

### **Problems with the measurement of outcomes in the North Thames study**

The North Thames study was undertaken prior to the study on which this PhD is based. The study is described more fully in chapter 3. What is important to highlight in this chapter is that it demonstrated that using a disease-specific quality of life approach did not produce useful or informative results. The open-ended interviews undertaken alongside the quality of life measurements suggested that some patients who attended PDC felt that their attendance had a positive impact on their quality of life. Furthermore, the pilot study for the economic evaluation component of the North Thames study also indicated that patients who attended PDC might have preferences for specific aspects of PDC. These preferences might differ from the reason given by the health care providers for why they attended. These findings

were speculative. The study was not designed to provide any insights into how strong patients' preferences were for particular aspects of the services, or whether different configurations of the service (which existed across the five centres in the study) might have an impact on how much they wanted to attend.

One argument for why the quantitative findings of the North Thames study did not show any change in quality of life for patients attending PDC is that there is no tangible change in quality of life that relates to attending PDC. Another is that the domains of quality of life that were assessed were not the relevant ones for PDC patients. It could be argued that since the palliative care quality of life instruments were developed in an inpatient context and concentrated on pain and symptom control, they reflected a more acute experience of disease than that experienced by patients while attending PDC. For patients in the North Thames study, pain and symptoms were usually well controlled in order for them to be able to attend a PDC centre. The qualitative analysis described the domains of quality of life that appeared to be important to patients who attended PDC. These were social ("getting out and meeting other people") psychological ("being around people in the same situation", "being able to talk about problems", "someone there when you need them"), and about the impact of specific therapies and interventions that lead to more relaxation and more physical and emotional well-being, ("feeling like a person again") (Goodwin et al 2002). These are not aspects of quality of life that can be easily described in words nor described in terms of specific, discrete domains of quality of life.

Following this argument, it might be the case that a quality of life instrument that captures more global dimensions of well-being would be a more appropriate instrument for capturing the important outcomes of a service such as PDC that aims to meet a wide set of health and welfare goals. A global quality of life instrument should in theory be sensitive to any changes in quality of life, not only those related to a specific disease or health state.

However, there is another argument arising from the evidence from the North Thames study that would suggest that a quality of life instrument is not the appropriate way of establishing the value of PDC services. If one of the objectives of PDC is to meet the needs of patients who attend, and since these needs may be very different depending on the individual and their circumstances, then a quality of life approach to evaluation may not be useful. This argues for a more radical departure from straightforward quality of life outcomes in economic evaluation. It requires a different class of evaluation methodologies that can incorporate more deliberately the preferences of individuals for particular aspects of PDC.

PDC is not one service and people use it in different ways for different needs. A health gain approach that assumes that PDC is one intervention may not be sensitive enough to detect this complexity. The results, equivocal or unequivocal, may not provide the type of information that is needed in order to establish what it is about the service that patients (or groups of patients) require in order to meet their different needs. There is very little evidence to support decisions about which aspects of PDC seem to be more important to patients (or groups of patients, or individuals in specific circumstances) and how strong patients' preferences are for specific components of the service.

Chapter 2 reviewed the arguments for a preference-based approach to establishing whether a policy should be supported. Cost-benefit analysis is the 'gold standard' approach to economic evaluation but has not been adopted widely in health care evaluation since the problems of measuring revealed preference are onerous and other methods of measuring welfare have been adequate as proxy measures of revealed preference (even where this has not usually been explicit). For interventions such as PDC, where quality of life outcomes are not well described and the relationship between PDC inputs and outcomes is not clear, it may be appropriate to re-examine whether a preference-based approach to measuring welfare is a more appropriate form of evaluation.

The following two sections of this chapter consider these broad arguments in more detail and review the theory behind global quality of life and preference-based approaches to measuring welfare.

### **SECTION 3. FROM DISEASE-SPECIFIC TO GENERIC MEASURES OF HEALTH-RELATED QUALITY OF LIFE**

#### **Adopting the EQ-5D instrument in the PDC study**

EQ-5D is a simple generic 5-item quality of life instrument designed to measure health outcome (EuroQoL Group 1990). The instrument produces a weighted score for each patient that, when combined with any additional life years gained, produces a composite measure of quality-adjusted life years that can be used as the measure of outcome in cost-utility analysis (Kind et al 1998). Five domains of health related quality of life are self-assessed by respondents: self care, usual activities, pain and discomfort, anxiety and depression. Respondents are asked to rate themselves in all five domains as either having 'no problems', 'some problems' or acute problems, described as "confined to bed", or "in extreme pain and discomfort," for example. In this way, 243 possible health states have been identified. A visual analogue scale (VAS) from 1 to 100 is presented and respondents are asked to rate themselves in terms of how good they feel (up to 100) or bad they feel (to



0) today. This score is not incorporated into the weighting score but provides additional information on patients' well-being. This has been interpreted by one of the original contributors to the EQ-5D project as an individual's level of 'morale' on any given day since it is a reflection of the overall well-being of a person at any given moment in time (Williams 2000). This interpretation is plausible because the value assigned by an individual to a VAS score is determined more by their mental or moral confidence, feelings of hope and optimism than a more objective assessment of their well-being. However, this hypothesis has not been explored empirically.

A major research activity in the EQ-5D project has been to identify the value that society places on each of these 243 identifiable health states. Studies to estimate population-level valuations (how much worse a health state is than full healthy life) has been undertaken in the UK (Kind et al 1998), and replicated in Europe (Badia et al 1999, Bjork and Norinder 1999) and worldwide (Nord et al 1993, Burstrom et al 2001, Devlin et al 2002, Tsuchiya 2002) and show similarity of scale values. Once population values of health states have been identified, a composite score is calculated for each health state. This score is the weighted adjustment that can be made to a full life year in order for quality of life to be incorporated into measurement of outcome.

Validating the methods for deriving QALY estimates is on-going and has produced some incongruous results. These suggest that the measurement of outcomes may be dependent on the approach adopted (Hollingworth et al 1995, Dolan et al 1996, Jenkinson et al 1997, Krabbe 1997, Glick et al 1999, Badia 1999). However QALY estimates using instruments such as the EQ-5D have been put forward as the gold standard approach to economic evaluation by The National Institute for Clinical Excellence (NICE 2001).

### **Advantages of EQ-5D over disease-specific quality of life instruments**

The EQ-5D has specific advantages over disease-specific measures of quality of life in the context of complex services such as PDC. First, the global measure of quality of life may measure benefits more effectively than measures that were designed for acute interventions. A global measure may be a more appropriate reflection of the overall importance of a service (or health and well-being in general) in patients' lives. Second, measuring health-related quality of life in more generic terms allows comparison across different kinds of interventions. Also, explicit values are assigned to health states in order that outcomes reflect subjective valuation rather than only objective measures of well-being.

## **Use of the EQ-5D questionnaire on people who have long-term illness or disabilities and older patients**

There has been some empirical evidence published that suggests that the experience of ill health has an important effect on health state valuation. The research found that patients with long-term illness valued health states significantly differently from healthy individuals (Badia 1998). Healthy individuals assigned some health states a negative score (implying states worse than death) while patients with long-term illness assigned positive scores to all health states.

Other empirical research among health care providers found that EQ-5D scores from the general population did not adequately describe the valuation of health of people with disabilities (Taylor et al 2001). The scores derived from health professionals who worked with patients with long-term illness were significantly higher than those of the general population. This research did not obtain valuations from the patients themselves for comparison. Variables such as years of experience and type of profession had a significant impact on scores, as did the type of disability described. The ability to perform usual activities had the most impact on valuations.

The effect of the experience of illness and age on health state valuations has also been explored. Empirical evidence has shown that those in poorer health generally give higher scores for health states (Dolan 1996). However, the scores for patients over 60 years have been found to be considerably lower than those based on values of people aged 18-59 (Dolan 2000). In this case, we could expect that younger patients would give higher scores for health states than older patients. No valuation project has provided estimates of EQ-5D valuation scores for patients with a life-limiting illness and who are no longer receiving active treatment. These valuations may differ from the general population since people at the end of their lives may value even small changes in quality of life that allows them to live their last months in relative peace and without pain. This is discussed again further on.

There has been work undertaken to establish the validity of the technique in elderly populations. Age has also been found to have a significant impact on the ability to complete the EQ-5D questionnaire (Brazier et al 1996, Coast 1998).

## **Would EQ-5D be sensitive to changes in quality of life in a PDC setting?**

The advantage of the EQ-5D outcome measure in economic evaluation is that it can be used to calculate quality adjusted life years. This would overcome some of the problems of previous studies that only considered clinical and disease-specific quality of life endpoints.

There are *a priori* reasons for suggesting that the EQ-5D instrument and the generation of QALYs may not be an appropriate instrument for measuring the outcomes of palliative care. These reasons relate to the issue of time and the issue of sensitivity of the instrument to perceiving changes in quality of life along the domains that are important to patients in this phase of illness. There are several issues to consider. The first is whether the instrument is sensitive enough to distinguish different states of health. If the instrument were insensitive, then it would not show changes in quality of life even when respondents report that a PDC intervention has made an important difference to their quality of life. Second, the valuation of health states should reflect the values of patients using PDC. As discussed earlier, there has been extensive work undertaken to identify population level valuations for all 243 health states but this work has not been undertaken for specific groups such as those at the end of life.

The value of an additional QALY is the same regardless of when it is lived. For palliative care interventions, there may be reasons why people value health state differently at the end of life. For example it might be more important to be able to take part in special activities with loved ones at the end of life than in other periods of one's life. People may be less (or more) frightened of pain or some discomfort at the end of life than at other times. An improvement in the quality of time at the end of life may be more valuable (even if it is only a few days) than the same absolute improvement earlier in life. This is because people may not expect to have any improvements in quality of life at this stage. Also they may highly value any additional quality of life in order to be able to undertake particular tasks or activities that are important to them at the end of their lives. This may also be extremely valuable to their family and loved ones. These individual preferences cannot be captured by a QALY approach without also undertaking an exercise to re-evaluate the QALY weightings for this group of patients. These ideas have not yet been fully explored in the QALY literature. This may be an avenue for future empirical investigation.

The valuation of health states may also vary between patients who use PDC services. In an editorial on palliative care it was suggested that patients experience their objectively similar levels of illness in subjectively different ways:

*"Some patients with minimal dysfunction are extremely dissatisfied while others seem quite able to tolerate severe impairment and may even feel fortunate to obtain therapy. Patients' perceptions of their illness are extremely variable and factors other than their disability come into their perception."* (Cella 1995)

Another important issue is that QALYs are generated by measuring additional years of life weighted by the quality of life in a particular health state. However good (or bad) a health state, the endpoint for the outcome is a time-based measure. Since the majority of patients who require palliative care are not expected to live long, the fact that the outcome measure is dominated in units of time is problematic. If an intervention that improves life expectancy by only a fraction (say from 6 months to a year) were to be compared with a palliative intervention that improves quality of life, the palliative intervention would have to have an impact on quality of life equivalent of 0.5 to be equivalent (say from a health state valued as 0.2 to one valued as 0.7). An assessment of how likely this might be in a PDC setting can illustrate this point. Table 4.1 shows the calculation for estimating quality of life weighting for each EQ-5D health profile. Full health is given the value of 1.0. Any state of health less than full health is estimated by subtracting from 1.0. The constant term, for any state less than perfect health is 0.081. If level 3 occurs in any domain, an additional parameter of 0.269 is subtracted. The values to be subtracted for level 2 and 3 for each domain are given in the table below.

**Table 4.1 Values for estimating weighting for each state of health. Values for the UK general population** (source: Kind, P., P. Dolan, C. Gudex, and A. Williams (1998) Variation in population health status: results from a United Kingdom questionnaire survey. *BMJ* 316, 736-41)

EQ-5D dimension	Level 2	Level 3
<i>Full health</i>	1.0	1.0
Mobility	-0.069	-0.314
Self care	-0.104	-0.214
Usual activity	-0.036	-0.094
Pain/ discomfort	-0.123	-0.386
Anxiety/ depression	-0.071	-0.236
	Constant for any level two or higher, subtract 0.081	Any level 3, subtract an additional 0.269

A hypothetical PDC patient at first attendance would be expected to be able to get out of bed (with help from a carer), be able to travel, be able to eat with others and socialise to some limited extent. Very few patients would be in either acute pain/ discomfort, unable to perform any usual activities, or be completely unable to wash and dress. They may have acute anxiety or depression (level 3) and have some problems with all other domains (level 2). This could be represented as a score of 2 2 2 2 3, calculated as a weighting of 0.082. Marginal improvements in this score, say in anxiety and depression (moving from level 3 to level 2) and in pain discomfort (level 2 to 1) and usual activity (level 2 to 1) could be represented as the profile 2 2 1 1 2, which has a weighting 0.675. This would represent a difference of about 0.6 QALYs if a patient was in this improved state for a year. If the patient

lives for only 6 months in this health state, then the intervention would have produced around 0.3 additional QALYs. This level of improvement might be expected in PDC, given the aims and objectives of the service. This is equivalent of an intervention that increases life expectancy (or reduces premature death) by about three and a half months.

The impact of this simple analysis demonstrates that interventions that increase length of life by only few months will dominate life improving interventions such as PDC. This argument does not consider the relative outcomes of interventions for this group (which may always be small) nor the likely number of individuals who might benefit from PDC (and represent a small benefit but for a large number of people). However, this example suggests that any benefits that do not increase length as well as quality of life will not fare well in analyses that compare outcomes across patient groups and health care settings.

Table 4.1 above shows that moving from level 3 to level 2 in the domain of 'pain and discomfort' represents a marginal change of 0.263. For 'usual activity', the difference is 0.058. This implies that improvements in pain and discomfort are more highly valued by the general UK population than improvements in performing usual activities. This makes intuitive sense. However, PDC is more likely to have an impact, given the nature of the intervention, on improvements in usual activities and anxiety and depression (marginal change from level 3 to 2 of 0.165) than in pain control. It would also require a larger sample size to detect significant differences in quality of life between groups than would be necessary if the expected change was in pain or mobility. This is potentially problematic given the challenges of recruitment and retention of subjects in palliative care research.

Second, people who attend PDC are not expected to regain full health and may not live for a full year. They would be expected to continue to have some problems and deteriorate until death. Since time in any health state might be measured in weeks or months, the detection of a significant change in QALYs would also require a large sample size to detect small differences. These differences in time lived in a particular health state may not register as significant in the research findings but, as described earlier, may be very important to the individuals experiencing an increase in number of months or weeks (or even days) of life or small improvement in quality of life in this period of their lives.

As there are only three levels of distress for each domain in EQ-5D, the likelihood is that patients in a palliative phase of illness may improve but continue to have 'some problems' in all the domains. This means that respondents would continue to tick the box 'some

problems' even though they may have marginally more or fewer problems. This distinction will not be picked up using this three-level approach.

Research undertaken prior to the development of EuroQoL (a precursor to EQ-5D) in the 1980s had suggested that QALY approach to assess the outcomes of long-term care for older people in the UK would be insensitive to changes in health status of older people in long term care (Donaldson et al 1988). It was argued that the dimensions that were appropriate for this patient group (disability and distress) would need to be included in a QALY instrument.

The extent to which these arguments are also relevant for palliative care patients has not yet been explored. The palliative care literature has focussed on domains of quality of life such as pain and symptom control and psychosocial aspects of living with a life-limiting illness (Massaro 2000). It has also broadened the debate about what constitutes health and quality of life by trying to conceptualise ephemeral aspects of quality of life as "existential health" (Doyle 1992, Bolmsjo 2002, Albinsson 2002) or "spiritual health" (Breitbart 2002, Nelson 2002). This has to do with finding hope and meaning in life at the end of life by those experiencing the illness as well as the people around them.

EQ-5D is a standardised instrument for use as a measure of health outcome designed to be applicable to a wide range of health conditions and treatments. Domains such as pain/symptom control and anxiety/depression that are relevant to palliative care are covered by the EQ-5D instrument, but it is clear that other domains such as existential or spiritual health are not. These domains are not clearly understood or even accepted as relevant domains of health-related quality of life (Warr 1996, Kaasa 2002). The EQ-5D instrument should not be criticised to something it did not set out to achieve, but the limitations of its use in palliative day care population are that it might not capture the domains of health-related quality of life (or quality of life more generally) that are important to those who attend a centre or are responsible for the care that is provided. However, the use of this instrument in a palliative care population has not been explored to date, except in one very recently published study of radiotherapy palliation (Van den Hout et al 2003). However, this intervention has more in common with mainstream health care than palliative day care.

An opportunity arose to use the EQ-5D questionnaire in a second study of PDC undertaken after the completion of the North Thames study in 2000. The methods used, and results from this study are presented in the next chapter.

## **SECTION 4. THE CHOICE EXPERIMENT APPROACH TO EVALUATING PALLIATIVE DAY CARE**

Modelling consumer preference has been a major activity in consumer research since the 1960s (Green and Srinivasan 1978, 1987). This section provides an overview of the theoretical foundations of stated preference choice experiments (CE) and presents arguments for why this approach may provide evidence of the value of PDC in a way that is useful to decision-makers. It involves a more in-depth methodological analysis than the discussion of the EQ-5D since it is a less established, more recent development in health economics and there is some controversy around its use in economic evaluation. Furthermore, choice experiments (CE) had not been tried out in frail elderly populations, patients in a palliative phase of illness or cancer patients when the design of this study was first underway (although there have been more recent studies published in this area which will be referred to later). For this reason it was necessary to consider its theoretical strengths and weaknesses and the appropriateness of its application to PDC research before attempting to use this methodology in a study of the views of potentially vulnerable people.

### **Definition of a choice experiment**

A choice experiment is a stated preference technique for establishing the importance of individual attributes in the overall utility of a good or service (Cave et al 1993, Ryan 1996). Stated preference methods are “a family of techniques which use individual respondents’ statements about their preferences in a set of .... options to estimate utility functions.” (Kroes and Sheldon 1988).

Ryan and Hughes (1997) have argued that choice experiments are another method of estimating an individual's utility in a way that is similar to standard gamble and time trade-off techniques. Unlike these techniques, choice experiments estimate the utility of particular attributes of an intervention and establish the relative importance of different attributes by estimating the marginal rate of substitution between them. Another argument for the choice experiment approach in evaluation is that it resembles “real life” choices and is therefore superior to other contingent valuation methods. The assumptions underlying choice experiments will be explored further on.

The approach assumes that a service (or product) can be defined in terms of a few important characteristics of that service (Ryan, Bate et al 2001). Also, it is assumed that when an individual makes the decision, it is based on trade-offs among these characteristics. An individual decides which characteristics (or attributes) are important to them and which are less important and makes a choice that will maximise his or her utility from that service.

All forms of the approach are concerned with the valuation of attributes of a good or a service and transforming respondents' subjective preferences for attributes of a service into numerical valuation of the attributes. These numerical valuations have different properties and can be interpreted in different ways depending on the approach and underlying theory. For example, marketing researchers have been concerned with disaggregated values of attributes and in designing commodities with optimum value attributes, that is, the highest overall value to customers (Green and Wind 1975). By contrast, economists have been mostly interested in the aggregate implications of multiattribute utility structures in terms of how they help to describe the aggregate demand function for a particular good or how the results might be used in cost-benefit analysis (Cave et al 1993). Public sector economists have interpreted this approach as being a way to understand and measure the utility of a good or service where a market value may not exist (Adamovicz 2002). In other words, it may be used as a proxy for measuring monetary value directly (i.e. by individuals' revealed behaviour in the marketplace).

Therefore this approach may be best suited to problems where the decision is based on attributes rather than the whole good or service. It can also be used where the research or policy problem is to value one or more attributes or where a specific combination of attributes does not currently exist in the market or new attributes are being considered (Green et al 1988, Haaijer and Wedel 2000). The method makes a number of strong assumptions about the functional form that can transform the attributes into additively separable utilities for individual attributes (also known as part-worths in the marketing literature) and sums the utility of individual attributes to arrive at the overall utility of a product.

### **How a choice experiment works**

There are different ways to design a choice experiment. In all approaches, respondents are asked to make trade-offs between scenarios that are presented to them that have different characteristics. A scenario is a simplified version of a product, a service, or outcome that can be described by its attributes. Respondents may be asked to choose between two or more pairs of scenarios. For each, say, pairwise choice, individuals select the option that, in their mind, would bring them the highest level of welfare. For example, a dessert might be described by its flavour, its colour, its size and its price. These attributes may be quantitative, such as time, or price. They may also be descriptive of a specific attribute, such as colour.

In a pairwise choice experiment, two scenarios representing the same product or service but describing different levels of attributes (higher price, different colours etc) are shown to the respondent. The respondent is then asked to choose which, overall, taking all attribute levels



into account, they would prefer. Some attributes are seen as positive (flavour, colour) some as negative (higher cost). Respondents must trade between positive and negative attributes in each description and decide which one, taking in all the attributes, they would choose or maximises their overall utility.

This task is repeated with different pairs of scenarios (or holding one scenario constant in each round), and asking the respondent to decide which they would choose each time. Each response (say, in a pairwise choice) for each respondent is entered into a regression model. The results estimate the relative importance of each attribute to the decision to choose a scenario. If the attribute is not important, the parameter in the model for that attribute will not be significant (for example, a p-value below the 0.01 significance level). If it is important to them, it will be shown to be significant at this level.

The relative size of the coefficients for each of the attributes is interpreted as the increased propensity to choose a scenario if the attribute is present (or at a higher level). It indicates how important, relative to the other attributes, the presence of a specific attribute is to a respondent's decision to choose a scenario. The signs on the coefficients in the probit models can be interpreted qualitatively to indicate the direction of the association between the explanatory variables and the chance of choosing a scenario. If the coefficient is positive, then the presence of the attribute is positively related to the respondent's decision to choose a scenario. In other words, respondents have a positive preference for the attribute. Clearly, the signs and their interpretation in a probit/logit model will depend on how the attribute is described in words in the choice experiment, that is, whether the attribute is described as a positive or negative contribution to overall welfare.

The marginal rate of substitution between attributes can be calculated to establish the relative importance of each attribute relative to other attributes (or how much a person would be willing to sacrifice of one attribute to have more of another). If one of the attributes is a financial charge or a price, then the marginal rate of substitution can be expressed in monetary terms. If other information on respondents is gathered, then it might be possible to identify whether different groups of people have different preferences.

As individuals are accustomed to making choices on a range of stimuli simultaneously in normal market transactions, the experiment is designed to mirror real choices that people make in their daily lives. In general, the attributes should be those that are most relevant to actual or potential consumers (Cattin and Wittink 1982). It can establish the relative

importance of different attributes based on the notion of opportunity cost, sacrifice, or benefit foregone in making decisions to maximise overall utility within resource constraints.

It is interesting to note that, in the marketing literature, the paired comparison approach is seen as the least efficient design in terms of the information obtained per unit of respondents' time (Green and Srinivasan 1987). The recommendation is that rank order approach will fare better in terms of predictive validity than direct paired comparison, the approach taken in most choice experiments. Nevertheless in the economics literature, the paired comparison or trade-off approach based on random utility theory (described further on) has emerged as the dominant approach to eliciting values for multi-attribute products because of its particular properties of constrained choice and adherence to the axioms of utility theory.

### **Interest in choice experiments CE for the evaluation of health services**

One of the principal advantages of the CE method is that it allows experimentation on data which may not exist in the real world. This means it has important advantages over other methods that use data from observation, either of outcomes, or revealed demand (Cave et al 1993). Where these data are difficult to collect, it allows information to be gathered where otherwise none might be available for decision-making purposes. Related to this, it also allows for the valuation of products or services that do not yet exist and for the valuation of particular (secondary) aspects of a product or service that might be "swamped" by other primary attributes (ibid. 1993). These properties make CE techniques potentially attractive for evaluating complex health services.

The conceptual framework appeals to health economists working on public policy issues because it is grounded in Lancaster's microeconomic theory that recognises that the utility of a good or a service can be decomposed into separable characteristics. One particular property of a choice experiment is that it can provide a large amount of information from a relatively small amount of data from respondents (Cattin and Wittink 1982). In this sense it is a highly efficient research tool. Another property of choice experiments is that they do not rely on comparative data, either over time, or between patient groups. This makes it an attractive method for research in contexts where there is a particular problem of patients being too frail to continue in a study, or dying before they can be followed up (McWhinney 1994, Jarvis et al 1996, Grande et al 2000a).

The interest in choice experiments in health economics literature appears to have been driven by specific areas of economic research. The first is as a way of valuing aspects of

care “beyond health outcomes” as a means of deriving willingness-to-pay valuations without using lengthy and cognitively challenging stated preference techniques (Ryan and Farrar 1995). The second is the potential use of technique is as a way of evaluating interventions and services that have multiple attributes or properties (Ryan, Scott et al 2001)

Recent reviews of the discrete choice literature in health economics have demonstrated the variation in study question and study design. A review of 31 choice experiment studies concluded that the methodology had been interpreted in different ways for different contexts but that methodological transparency and adherence to theory had been improved over time. What was lacking in many of these papers (and consequently from the review) was a detailed reasoning behind the different approaches to undertaking choice experiments and therefore why there was such heterogeneity in study design (Ryan and Gerard 2001). Furthermore, the impact of variation in design has not yet been fully explored (Ryan, Scott et al. 2001; Ryan, Bates et al 2001).

### **The economic theory underlying choice experiments**

It has been argued that the choice experiment approach originated in the literature of mathematical psychology in the early 1920s. It was then was adopted by marketing researchers, then by geographers, sociologists and planning analysts (Huber 1987). These first approaches were not based on any explicit theory of value, nor were they an attempt to measure utility or welfare in the way that it is understood in microeconomic theory. The use of the terms utility and part-worths in the marketing literature did not have explicit economic meanings but were terms coined to express the measure of value of specific attributes of a marketed good (Green and Wind 1975).

In their seminal paper, Luce and Tukey pioneered the trade-off approach whereby respondents react to multiple stimuli simultaneously (Luce and Tukey 1964). They were concerned with the joint effect of two or more independent variables on the ordering of a dependent variable (Green and Rao 1971). It was also recognised that such an approach could have some uses in cost-benefit analysis in assessing the trade-offs between different policies (Green and Rao *ibid.*).

Economists have considered adherence to utility theory to be of great significance, precisely because of the desire to be able to interpret utility in choice experiments the same way as it might be interpreted in cost-benefit analysis (Small and Rosen 1981). Chapter 2 considered the fundamental axioms of expected utility theory. The violation or adherence to the axioms of this theory in choice experiments has been an area of concern. A programme of research

in health economics is currently being undertaken to explore the underlying axioms of choice experiments and the consequences of violating these axioms under different conditions. The importance of this work is that it sheds light on the extent to which choice experiments reveal underlying (or latent) utility. If the axioms fail (and the assumption about human behaviour does not reflect the real world of decision-making) then it is more likely that the approach is akin to a more pragmatic decision-making approach. The outcome of this work will determine the extent to which the findings of a choice experiment in PDC are valid in reflecting the monetary value of specific attributes of PDC. Another equally important contribution is to provide evidence of the relative value of attributes of PDC for decision-makers.

### **Lancaster's theory of value**

Classic demand theory is based on assumptions about human psychology. Value is determined by how individuals satisfy their wants and is concerned with the factors that shape preferences for commodities (Rabin 1998). Until Lancaster's theory of demand was developed, little progress had been made in understanding how the properties of the goods themselves affect demand or in predicting how changes in the goods would affect individuals' preferences for them. Lancaster's contribution was to consider the properties of the goods themselves and identify the objective characteristics of goods that are relevant to choice (Lancaster 1966;1971). His fundamental proposition was that all goods have characteristics that are objective and finite and satisfy human wants. He also proposed that goods or services are demanded because they contain certain characteristics that contribute to utility. Therefore individuals' demand for commodities is a derived demand for the characteristics they contain. Lancaster made this point succinctly:

*"[T]he good per se does not give rise to utility to a consumer; it possesses characteristics, and these characteristics give rise to utility."* (Lancaster 1966)

Second, people possess preferences for different bundles of characteristics. This is the subjective, personal aspect of demand. Since Lancaster recognised that people possess different preferences for different collections or bundles of goods, these preferences are indirect or "derived" since the goods are only required to produce these characteristics. The "characteristics" models of demand separates the aspects of demand that are universal (depending on objective characteristics of goods) and those that are specific to the preferences of the individual.

Lancaster was making the point that demand is made up of two elements, people and things. He argued that microeconomic demand theory had not made use of information about “things” (or commodities or products). It did not trace effects of changes in the physical properties of goods, so it took no account of information that is readily available. While not all goods have the same properties to all people, some properties are clearly universal. Also, goods have many properties, but only some of them are relevant to choice. Those that are relevant to choice Lancaster has referred to as “characteristics”, that is objective properties of things relevant to choices by people.

This “characteristics” theory of value recognises that the utility of a good or service can be decomposed into separate utilities or attributes and allows examination of preferences for different goods and services by their constituent parts so that:

$g_k$  can be described by its specific attributes  $x_1, x_2, \dots, x_{ij}$

where  $g$  is a specific good with  $K$  attributes,  $i$  is all individuals  $1, \dots, I$  and  $j$  represents the levels of attributes  $1, \dots, J$

Multi-attribute utility theory on which choice experimentation stems from Lancaster’s characteristics theory of value by postulating that the utility of a good stems not only from characteristics but from preferences for specific *levels* of these characteristics (Herrman 2000, Green and Kreiger 1991). In experimental research, the levels of attributes that represent a specific good are assigned a numerical value such that the sum of all attributes at specific levels represents the total utility to an individual undertaking the experimental task. The role of the researcher is to find the function that aggregates the attribute utilities to an overall utility ( $U$ ) for the good  $g$  with  $k$  attributes and the good  $g$  with  $m$  attributes, so that

$$U(g_k) \sim U(g_m), \forall k, m$$

holds only if good  $g_k$  is preferred to  $g_m$  (Fishburn 1988, Green and Kreiger 1993).

This approach still makes the important assumption that people have well-defined and stable preferences for bundles of characteristics to meet their needs/wants. (Lancaster 1966, 1971).

### **Lancaster's influence on attribute based choice experiments**

Choice experiment theory has its roots in Lancasterian demand theory that relates demand to the characteristics of goods rather than to the characteristics or circumstances of the people who demand the goods. In Lancasterian theory, all attributes or bundles of characteristics can be combined in an infinite number of ways in order to determine the point at which utility can be optimised along a perfect continuum of attributes. Choice experiments attempt to mirror this choice continuum by expressing the relative value of attributes in terms of marginal rates of substitution, which assumes infinite substitutability.

In the real world, there are constraints to individuals maximising their utility since not all combinations of characteristics of products are available. The ability of individuals to make choices and trade-offs to maximise their utility in every consumption decision in every moment is constrained by time, cognitive overload and their willingness to undertake such a laborious task for all decisions in everyday life. There are also potentially millions of interaction effects (i.e. individuals' preferences for one bundle of goods being influenced by the availability and preferences for other bundles of goods) that would be impossible to take into consideration for all choices.

Choice experiments therefore operate *as if* the combinations of attributes and attribute levels are the only inputs into an individual's utility function for a particular product at a given time and the usual *ceteris paribus* rules apply. Respondents are assumed to be constantly seeking to maximise their own utility, but are only offered a limited number of attributes, attribute levels and trade-offs in order to do so. The results of the choice experiment are therefore limited in reflecting reality (and utility maximisation) and the importance of this limitation is determined by how closely the experiment reflects the real trade-offs and choices presented to individuals in the real world.

### **Random utility theory**

*"Generally speaking, there can be no valid measurement without underlying theory of the behaviour of the numbers which result from measurement"* (Louviere et al 2000, pp25).

Choice modelling is based on random utility theory (RUT). The development of random utility theory (McFadden 1973, Hanemann 1994) and statistical design theory has been fundamental in the development of the methods of choice experiments. Together they provide a way of modelling the choice experiment decision process (Ryan and Gerard 2001) and estimating the importance of characteristics of a product in a utility function. Random utility theory provides a unified theoretical framework to develop models to account for real

market choices (McFadden 1973, Louviere et al 2000). Using this theory, behaviour can be modelled in order to formulate probabilistic discrete choice models that can be estimated directly from choice experiments (Louviere *ibid.*). The RUT framework has placed choice experiments into the mainstream of microeconomic theory.

Random utility theory recognises the consumer's true utility for a particular good or service is unobservable and can only be estimated by modelling behavioural responses. These models have both deterministic and stochastic components (Manski 1977). Mathematical analysis that linked choice experiments and microeconomic theory through the application of random utility was undertaken in the 1980s (Beggs, Cardell et al 1981, Train 1986). This analysis relates choice experiments with the larger family of probabilistic modelling in econometrics.

In random utility theory, while the individual respondent knows the nature of his or her preferences, the researcher does not. One of the main contributions of McFadden's work was the recognition that choice behaviour is stochastic when seen from the vantage point of the analyst/observer (Louviere 1994). Data from groups of individuals leads to more randomness and variation. It recognised the impossibility of accounting for all differences in individuals. Fluctuations in behaviour within individuals (who may not always appear to choose what they like best) can be modelled using a random component in the utility function, such that;

$$U_i = V_i + \theta_i$$

where  $U_i$  is the unobservable but true utility for individual  $i$ ,  $V_i$  the observable component of utility and  $\theta_i$  is the random, unobservable component.  $V$  is the proportion of the variance in a choice that can be explained and  $\theta$  is the proportion that cannot.

The random term  $\theta$ , or error term, is itself made up of two components:

$$\theta = v_i + \varepsilon_i$$

where  $v_i$  is the constant term specific to the alternative and  $\varepsilon_i$  is the random error term that is independently and identically Gumbel distributed over all alternatives (for example, in the logit model) or not independent or identically distributed for normal random variables (for example, probit models). The decision-making process within a choice experiment can be interpreted as an individual making a comparison of, say, two unobservable (to the

researcher) utility functions. The respondent is assumed to choose the option that leads to the maximisation of their utility at a given time. Since all that is observed is the final choice the respondent makes, all the factors that contribute to that choice cannot be known, or modelled. An error term that represents the unknown factors is included in the model to take account of the fact that the direct utility function cannot be determined.

This approach has led to a general framework for understanding and modelling many types of human behaviour. Utilities are assumed to have a random component, which mirrors the fact that the researcher is unable to understand or model perfectly all influences on choice. The analyst can, however, specify the probability that an individual will choose a particular option from a set of competing options (Louviere et al 2000).

This method of estimating utility makes the economic approach in choice experiments qualitatively different from the assumed functional form in the marketing literature where no error term is included (Van der Pol and Ryan 1995). It allows all the unobservable unknowable factors that affect utility and choice to be incorporated in the model. It represents the fact that underlying or latent utility can never be observed or known, but only estimated from observations of human behaviour under constraints.

In a choice experiment, faced with a choice of two scenarios, it is assumed that a respondent chooses the scenario that leads to the highest level of utility, so that

$$U_{iq}(A) = v_{iq}(A) + \varepsilon_{iq}$$

where  $U_{iq}(A)$  represents the unobservable, true indirect utility function of individual  $q$  for the good/service/intervention  $i$  with attributes  $A$ , and  $v_{iq}(A)$  represents the measurable component of the utility function estimated empirically, with  $\varepsilon_{iq}$  representing the unobservable factors.

### **Axioms underlying choice experiments**

Outside the economics literature, where a lot of applications of conjoint measurement and choice experiments have been undertaken, it was recognised that there were important general limitations to these approaches. These were specified in the first review of conjoint analysis techniques published in the 1970s (Green and Wind 1975). It was accepted that some goods and services might contain utilities that are not adequately captured by these techniques, especially where too little is known about either the product or the interactions between the utility of attributes within a product. Twenty years later, choice experiments



have been described as a “*double edged sword*” (Louviere 1994) in that “*the predictive power may come at the expense of real understanding. They may fit response data well but be incorrect and misleading*” (Louviere, *ibid.*).

The methods of choice modelling make a number of strong assumptions about human behaviour. Interest in the health economics literature is now focused on the assumptions or axioms in economic theory, rather than the theory of numbers, that underlie choice experiments and the extent to which choice experiments disclose underlying or latent utility (San Miguel and Ryan 2002a, San Miguel and Ryan 2002b, Ryan and San Miguel 2002, Cairns et al 2002, Amaya et al 2002). It is important to address some of these issues and whether they relate to choice experiments in particular or contingent valuation methods (and the neoclassical theory of demand) more generally. Understanding these core assumptions and how they are adhered to or violated will help in interpreting the results of the empirical work presented in the thesis. Some of this work is still being developed and, consequently, has not yet been published.

### **Importance of considering economic axioms in choice experiments**

Microeconomic demand theory is based on assumptions about behaviour and how it can be described in quantitative terms. Some of these assumptions apply to neoclassical theory of demand generally, some are specific to stated preference methods of eliciting values of benefits and some to the choice experiment approach. Research has been undertaken to assess the validity of these assumptions, and the extent to which these assumptions must hold in choice experiments in order to interpret the results as the values or utilities of attributes. This research is also trying to establish the significance of violations to these assumptions in terms of how the results might be interpreted. The violation of some assumptions may be more importance than others.

Health economists working on choice experiments have contributed to this field by undertaking a number of empirical investigations into the properties of economic axioms alongside collection of data on choice values. The evidence will be reviewed here in order to investigate how likely it is that the PDC empirical study can establish utility values for attributes of PDC, and the assumptions about the data that these findings must rest on. Some assumptions may be violated in some contexts but not in others. For example, a choice experiment where respondents may have experience of the attributes of a product or service (and know the consequences of their consumption, and their individual demand threshold for that product) may lead to more consistent choices than choice experiment

where respondents are unfamiliar with the product. The importance of these axioms in the context of research into complex services will be investigated.

### **Axioms of neoclassical demand theory : Rationality and choice**

It has been argued that the stated preference choice experiment approach to conjoint analysis (as compared with ranking and rating approaches) is superior because individuals are used to making choice decisions in their daily lives, whereas ranking or rating exercises are not as common in real life (Ryan 1999). Therefore it is seen as self-evident that respondents will find choice exercises easier to understand and respond to, and presumably therefore will give more valid answers. The limited empirical evidence available has reported high test-retest reliability in choice experiments, suggesting that respondents are consistent over different time points in their preferences, although this evidence is from one study only (Bryan et al 2000). Tests of reliability cannot demonstrate whether choice experiments are valid, that is, measuring what are designed to measure. In this case validity relates to the ability of the experiment to reveal respondents' true relative preferences for different attributes.

The validity of choice experiments as consistent with the economic theory depends on whether the observed preference meets the economic axioms of rationality. The concept of rationality is a theoretical underpinning of neoclassical demand theory. Neoclassical demand theory assumes that consumers are utility maximisers and will choose to consume goods and services that promote this goal, subject to income constraints. Individuals' preferences are also complete (without indecision) and stable in the short term. Individuals are introspective and utility maximisers. Within this framework, they make decisions that maximise utility and their preferences are revealed through their consumption.

### **Trading between attributes**

Another one of the fundamental axioms of utility theory is that people have complete and stable preferences. This that means that they have well formed preferences regarding the attributes of a product or service. It also assumes that people are willing and able to trade the attributes in any valuation task to maximise their welfare. This also assumes that there is always a level of one attribute that can compensate for deterioration on the level of another attribute. This has been termed the axiom of continuity. There is an implicit assumption that people make decisions, and engage in compensatory decision-making in a way that represents their true preferences (and therefore reflects their underlying utility).

Preferences for products, or attributes that make up a product are assumed to be continuous. This means that for every individual, there exists a price or marginal rate of substitution whereby each product is tradable with any other which can be represented on continuous indifference maps. Individuals must be prepared to trade between attributes to maximise their welfare. No single attribute can be dominant over all others (for which individuals are not prepared to trade at any level of compensation with other attributes). Such a pattern of preferences, where an individual would not trade no matter how much utility in other attributes was sacrificed for that attribute level, would be interpreted as irrational since preference for that attribute would be infinite.

Other tests of rationality also have to be met for choice experiments. Some are easier to test for than others (Gerard 2002). The straightforward tests of rationality are reported in box 4.2 below:

#### **Box 4.2 Concepts of rationality and how they might be tested in choice experiments**

##### **Completeness of preferences**

Test: Individuals should be able to state their preferences for a product. The consequences of consumption are known to the respondent, or can be formed (for a new product) if adequate descriptive information is given.

##### **Independence of preferences**

Test: The preference ordering within one attribute should not depend on the level of another attribute.

##### **Continuity and stability of preferences in this short run**

Test: Re-run the same set of pairwise choices with the same respondents at a different time point in the near future, *ceteris paribus*.

##### **Transitivity of preferences**

Simple test: If scenario A is preferred to B and scenario B is preferred to C, then individuals must prefer scenario A to C.

If these assumptions are satisfied then there exists an additive solution such that the utility of a product is equal to the sum of utility of all the attributes,

$$U(A,B...n) = U(A) + U(B) ..... U(n)$$

where U is the sum of all utility from a product and A, B .....n are attributes of the product.

There is evidence from the psychology and economic evaluation literature that these assumptions may not be valid. Evidence has been produced that shows clear reversal of preferences in studies of how individuals choose between two gambles, depending on how the gamble is presented (Lichtenstein and Slovic 1971). Other evidence has suggested that

preferences may not be stable or complete. For example Shiell and colleagues found that the majority of participants (in an admittedly small study of college staff and students) demonstrated unstable preferences on repeat presentation of a standard gamble task. This suggested that individuals were reflecting on their responses over time and changing them (Shiell et al 2000).

Evidence from the environmental economics literature has demonstrated that people's estimates of value do not reflect the magnitude of the commodity. One study showed that the maximum willingness to pay for environmental preservation of a lake in one area of Canada was similar to people's valuation of environmental protection of all lakes in Canada, describing this as a "purchase of moral satisfaction" (Kahneman and Knetsch 1992).

Transportation and environmental economics has been at the forefront of the theoretical development of more complex models for accurately predicting choices. If, for example, individuals exhibit non-compensatory decision-making, then this will severely restrict the extent to which in practice overall preferences between options (scenarios) can be established by CE methods (Swait 2001). Swait has argued for an extension of the neoclassical theory of consumer behaviour to incorporate 'cut-offs' which can significantly improve the choice prediction in stated preference choice tasks. His argument is based on insights that the normative economic model of the fully rational, fully informed, utility maximising decision-maker is restrictive and does not explain fully the complexity of human decision-making.

Swait's argument builds on the work of others in psychology, marketing and economics that individuals are in reality information processors with only limited capabilities and resources to make decisions. People are "cognitive misers" (Swait *ibid.* pp 905) and adapt the amount of effort they invest in making a good decision to their context and resources. Swait proposes that individuals adopt 'cut-offs' as a means to reduce the burden of the decision by eliminating alternatives in a choice set that do not meet specific requirements for which the individual will not trade. These cut-offs are psychological boundaries of particular attributes within which the respondents will restrict themselves. An example of such a boundary might be that a respondent will never choose an option with a price higher than £X. Non-compensatory decision-making means that a respondent is not willing to allow compensation between different criteria in a decision. Swait has shown that if models that incorporate such attribute cut-offs produce a better theoretical model and a sizeable improvement of the predictability of the CE model.

For discrete cut-offs the choice model can incorporate dummy variables that reflect whether the requirement is violated in a given alternative in a choice set. This dummy variable is included for all attributes that are outside a given cut-off. For this to be operationalised, the individual must provide his or her own cut-offs to the researchers. These can be decided *ex ante* with the danger that these preconceived cut-offs are not well formulated but will influence the respondents decisions in the experiment as people wish to be consistent in their preferences, or *ex poste* in which case there is a danger that the answers given in the experiment will influence the cut-offs rather than reflect the individuals true cut-offs or preferences in the real world. There have been difficulties in implementing choice models based on two-stage choice processes (for screening the attributes for cut-offs and second, making the final choice) as it can increase the number of choice sets exponentially (Swait and Ben Akiva 1987, in Swait 2001) but the approach proposed by Swait that includes dummy variables for cut-offs can capture both stages in the model parameters and so reduce the number of choice sets required to estimate the model.

Research in environmental economics has experimented with designs that can reflect the intensity of preference (Johnson et al 2000, Swallow et al 2000, reported in Adamovicz 2002). This allows examination of a subset of respondents who are confident about their responses. Adamovicz suggests that in the contingent valuation literature, this approach has improved consistency with actual choices. But there has been too little empirical examination of the effect of the additional burden of these approaches on respondents' ability to complete the choice tasks.

Some of the emerging issues in the choice modelling literature are whether the context of the experiment has an effect on the estimated parameters of the regression models (Adamovicz, 2002). Adamovicz argues that empirical evidence has shown that the regression model estimated from a choice experiment can change significantly with relatively small changes in context. Theoretical analysis of the importance of specific contextual issues (such as social interdependence of decisions, the complexity of decision, or the heterogeneity of preferences) is an emerging field in choice experiment research. It is likely that the consequences of this work will have an effect on how the results of simple choice experiments can be interpreted. Adamovicz argues that,

*"The critical issue is the extent to which the assumptions of the choice modelling approach can be satisfied and whether there is sufficient data to support the estimators..... it is becoming clear that the amount of data with many replications per individual may be necessary for the use of more advanced models."* (Adamovicz *ibid.*).

The more complex the nature of the inquiry into consumer choice, the more complex the choice experiment model may have to become, not only in the number of choices that are presented to respondents but in the analysis of those responses. However, Louviere (2001) has argued that any method of choosing choice sets would, given enough data, lead to good quality model estimates. Research effort has been concentrated in understanding and developing methods for estimating the regression models to fit the data: *"far less is known about the model parameters estimates, particularly their statistical efficiency"* (Louviere, *ibid.* pp19).

The debate outlined in the next section discusses the extent to which, given the restrictions of research design, choice experiments can claim to measure utility. It suggests that the violation of even the most lenient of the criteria for rationality (stability of preferences in the short run, within the same choice) means that it is difficult to interpret the findings of empirical research as measuring utilities.

### **Empirical investigation of the axioms in the health economics choice experiment literature**

Empirical investigation is still limited but it is addressing the validity of choice experiments beyond the very simple tests of rationality. The axioms have been examined with real data from choice experiments. At the same time, theoretical understanding has advanced to explore ways in which some more minor axioms might be violated but the overall rationality of an individual's responses remains intact.

San Miguel has undertaken work to suggest that individuals do not follow the decision process assumed by utility theory (San Miguel 2000b). The results are not sufficiently conclusive to enable rejection of utility theory but they raise questions about the need to explore its applicability in representing individuals' behaviour. Amaya and colleagues are examining whether individuals adopt compensatory heuristics in their responses to choice experiments. Heuristics are cognitive strategies or shortcuts that operate to simplify people's data processing when responding to a task (Kahneman et al 1982). There is a suggestion that individuals adopt simple decision-making heuristics in choice experiments that may be in direct conflict with the axiom of continuity of preferences (Amaya et al 2002).

Simplifying decision heuristics may exist such as 'Take the Best', where respondents consider one attribute at a time in the comparison of a choice set. If the first attribute can distinguish between the two options the two for the individual, then the decision is made on that first attribute (Gigerenzer et al 1999). Regression models using the 'Take the Best' rule

have been shown to be as accurate at predicting decision as using all the information presented to respondents (Gigerenzer et al *ibid.*). The authors consider these heuristics to be more psychologically plausible and a reason why the strict axioms may fail, but respondents are still providing rational responses.

Work by Ryan and colleagues is examining whether it is possible to demonstrate rationality without an understanding of the cognitive processes that a respondent uses to make a choice (Ryan, Hughes et al 2002). They have employed qualitative research techniques to examine the reasons given by respondents about why they make the choices they do in choice experiment contexts. They found that 30% of the choices that might be considered “irrational” could be explained by respondents not understanding the choice task (i.e. by making price-based not value-based choices). This suggests that respondents may in fact be making what to them are rational decisions although they violate the axioms of rationality. For example, a respondent may consider the cost of an intervention to them rather than the value of that intervention to them in terms of willingness-to-pay. There is also evidence from this work that people will employ simplifying decision-making heuristics in choice experiments contexts and consider only the minimum amount of information necessary to make the decisions where not all attributes are considered.

There have been two studies undertaken to explore Sen’s expansion and contraction properties in choice experiments. These are tests of rationality to explore whether respondents are making a range of decisions that are consistent when the questions are presented in different ways. The expansion property, for example, tests whether the choices made by respondents are consistent (rational) when choice sets are expanded or contracted. If a respondent makes a preference ordering between a pair of scenarios, this preference ordering should not change by the addition of a third scenario added to the choice set.

The first study, by San Miguel and colleagues showed that these tests were hard to satisfy compared with simple dominance and transitivity tests (San Miguel and Ryan 2002). Recently Gerard and colleagues have begun to investigate whether more irrational responses were detected when stricter tests of irrationality were applied (Gerard 2002). In the preliminary findings of this unpublished work there has been some support for the hypothesis that more stringent tests of rationality led to higher proportions failing the test. The consequence of this work is to challenge the notion that respondents do not act as if they are rational individuals when more stringent tests of rationality are applied. It is not enough to apply more simple tests and conclude from this that individuals are acting

rationally when faced with choice sets. It adds to the empirical evidence that utility estimates derived from empirical research may not always meet the strict normative axioms set out in theory.

The investigation of axioms is finding different results. On the one hand individual respondents' choices can be better identified as rational once complex decision-making heuristics are better understood by researchers. On the other, more respondents fail the stricter tests of rationality (Sen's test) when these stricter tests are applied.

There is some debate about whether testing for dominance is a relevant test in the context of a choice experiment. Dominant preferences are preferences for attributes that individuals are not willing to trade for any level of some other attributes. Since a foundational axiom of discrete choice analysis is that individuals trade to maximise their utility, the individual who expresses no willingness to trade regardless of the value of the alternative is perceived to have irrational preferences.

One approach to empirical analysis has been to class dominant preferences as irrational since respondents show no willingness to trade one attribute for any quantity of another in compensation. Therefore these respondents were excluded from the analysis (Ryan and Hughes 1997; Ryan et al 1998, Ratcliffe and Buxton 1999; Jan et al 2000). However, this position has been challenged. From a 'public policy' perspective it has been argued that since policy-making has its foundations in the democratic process, and some people do have irrational responses in choice situations, this should not exclude them from analysis (Bryan et al 1998). If these irrational responses are random then they should not have an important effect on the results. The qualitative evidence described earlier also showed that it is not possible for researchers to make *ex post* decisions about which patients are exhibiting dominant preferences (Ryan and Hughes 2002). There may be good reasons why people have what seems to be on the surface an irrational view even if these are not evident to the researcher who only ever has partial knowledge.

### **What is left of utility?**

*"The operation failed, but the patient thrived"* (Huber 1979)

This quote by Huber represents the phenomenon that despite widespread violation of the axioms the predictive value of choice experiments is quite strong. Utilities derived in this way may be useful despite these violations.



The value of the CE approach is that, unlike more straightforward ranking or rating exercises developed in the stated preference framework, it requires individuals to make decisions that require judgments to be made when choice is constrained. It is an improvement over direct elicitation techniques because it keeps respondents from seeing every attribute as important. Respondents are required to make decisions about what they would sacrifice in order to obtain a particular level of an attribute, and this provides some quantitative, comparative measure of its contribution to welfare, in the context where this cannot be measured directly.

The extent to which choice experiments are measuring utility is determined by the adherence to the axioms of rationality. The empirical evidence presented in this chapter and in chapter 2, suggests that individuals do not always act as if guided by rational decision-making. The question is the extent to which this matters in the interpretation of the results. It has also been suggested earlier in chapter 2 that utility estimation should be interpreted as a metaphor for decision-making (Nease 1996), and that adherence to the axioms is not necessary for the findings to be insightful.

There is evidence that decisions in the real marketplace are made on remarkably few dimensions (Olshavsky and Granbois 1979). The value of choice experiments is that they force people to make trade-offs that are similar to those of the market. It is analogous to a market choice and it requires judgements to be made when attributes conflict. It is an improvement on direct elicitation in that it keeps respondents from seeing everything as important. Even if choice experiments do not reflect utility directly, this process of systematically assessing how individuals value attributes of a product or service is still potentially useful information for policy-makers.

## **SECTION 5. THE METHODOLOGY OF CHOICE EXPERIMENTS**

The key stages of a choice experiment are introduced in this section, as described in the health economic literature published in the last decade. These studies build on methodological work developed in the environmental economics literature and transport economics. The health economics literature has contributed to the methodological development of the approach, especially in underlying economic axioms of choice experiments and the extent to which they hold true under empirical investigation described earlier. This has been an important contribution to the literature as it has an impact on the interpretation of the results of the empirical work undertaken in PDC.

### **How a choice experiment is constructed**

There are specific stages to designing a choice experiment and these have been well described in the literature and a recent review of the use of these methods in health economic evaluation (Ryan, Scott et al 2001). The main approach has not changed but there have been subtly different approaches described in the empirical studies. This reflects both the fact that this is an emerging field in health economic methodology, and that there are different ways of addressing the problem. An approach that has been widely followed in the health economist literature and which has acted as a type of blueprint for this kind of research is followed here (Ryan, Scott et al 2001).

There have been other approaches to choice set design suggested. One of these involves randomly selecting pairs of scenarios for respondents to consider. This has the advantage that each possible pair has the same probability of being compared by an individual (Adamovicz 2002). The disadvantage is that a large sample of respondents is required to take part in the study in order for there to be an adequate number of observations for each pair of scenarios. This is rare for choice experiments, and this approach has not been adopted in studies in health economics to date.

### **Establishing the attributes in the choice experiment**

From Lancasterian theory it is postulated that the characteristics of a good or service must be those that have an effect on a person's desire for that good or service (Lancaster 1971). If there is evidence from qualitative research that describes the important attributes of a particular service, this can be used to design the study. Equally, literature reviews or previous studies may provide data on the attributes and how to assign the appropriate levels of the attributes to the scenarios. There has been some discussion in the choice modelling literature on how these choices of levels then impact on the results. The consensus is that it is necessary to describe how and why the particular attributes and levels were chosen in order for others to be able to judge their appropriateness (Ryan 2000). Pilot testing will help to determine whether the choice of attributes is described in such a way as to be understood by the respondents.

### **Assigning levels to the characteristics**

Differences in levels need to be sufficiently wide in order to represent important differences in utility to an individual. As described earlier, the level of an attribute may be quantifiable and continuous (for example distance from clinic, where one mile is half as long as two miles), or descriptive (where qualitative differences in levels cannot be arithmetically

measured). Descriptive attributes are more difficult to infer as one person's interpretation may be different from another's. For example, for the "level of discomfort after the operation" the levels might be labelled: mild, moderate and severe, which may be understood by individuals differently. The importance for people to be able to discriminate between two choices is that they are both "plausible and actionable" (Ryan 2000).

Attribute levels can also be continuous (such as time or distance) or intervally scaled (such as the number of days per week that a clinic is open). Binary variables (for example if an attribute is either present or absent) can be represented by dummy variables in the regression analysis. This will be explored in more detail further on.

To minimise potential information overload, Thomas (1979, quoted in Huber 2000) suggests no more than five attributes should be included in a survey. Malhotra argues that respondents are capable of processing ten attributes without excessive strain (Malhotra 1982). Green and Srinivasan (1978) recommend that no more than six attributes should be used.

### **Number of choice sets**

The final product of the choice experiment is the estimation of a probabilistic regression model. Each of the coefficients or parameters of the model represents an attribute and shows the influence of that attribute on an individual's likelihood to choose the scenario. There have to be a sufficient number of choices presented to respondents so that these coefficients can be estimable to an acceptable degree of significance. One of the important properties of choice experiments is that it should be possible to estimate the coefficients without having to present all possible combinations of levels of attributes to all respondents. These can be calculated from a pairwise choice of only a fraction of the possible scenarios, using, for example, an orthogonal design.

An experimental study that incorporated all combinations of attributes, even for a fairly simple study, could be very large indeed (five attributes with three levels would lead to 243 possible scenarios). Since a full factorial design is usually not required to estimate the model coefficients, a method of reducing this cognitive burden on respondents is used. Empirical investigation of the number of choice sets found decreasing reliability with increasing numbers of properties (Acito 1979). The restriction is that the number of profiles must exceed the number of factors to allow for error degrees of freedom (Jones 2001).

A common approach to reduce the number of scenarios has been to use a fractional factorial design. Fractional factorial designs might be adopted when presenting all alternatives would be too time-consuming, cost too much, or might fatigue the respondents, thereby possibly invalidating responses (Kanninen 2000). A fractional factorial design is a subset of all possible combinations of attributes used in the experiment. One type of sub-set is an orthogonal array. It design allows estimation of the relative value of individual attributes but assumes that interaction between attributes is negligible. Although decision-making by individuals does incorporate interactions between different attributes of a choice, it may not be necessary from a practical point of view to include these in a choice model. It is reported that main effects typically account for around 70-90% of explained variance in choice models, with two-way effects accounting for another 5-15%. More complex interactions in decision-making probably account for around 5-10 only (Louviere, Hensher et al 2000). In an orthogonal array, there is no collinearity between attributes so that the probability of one level of attribute appearing in a choice set is not associated with the appearance of another level of another attribute.

The orthogonal design produces profiles (or scenarios) of combinations of attributes that have statistical properties which determine what utility specifications can be estimated from the response data, and with what efficiency. Virtually all orthogonal designs in health economic applications have used main effects plans (assuming no interaction between attributes) which is a potential limitation in terms of reflection of the complexity of interactions in real life choice situations. Despite this limitation, main effects designs are common with choice experiments because they do not require large sample sizes (Kanninen 2002).

The complexity in choice experiments may be similar or different to the complexity faced in real market decisions (Adamovicz 2002). The advantage of the fractional factorial design where not all combinations of levels of attributes are evaluated by the respondent is that the model can predict the equations for those combinations of attribute levels that subjects did not evaluate directly.

The analysis of choice data requires more complex analytical techniques than straightforward ranking or rating methods. Recent developments of suitable analytical procedures such as logit (logistical probability unit) and probit analysis and the use of software packages to execute these techniques have meant that CE methods have become more widely used (Cave et al 1993, Ryan, Scott et al 2001). Computer programmes can select a set of scenarios that are orthogonal and level balanced (each level of an attribute appears proportionately the same number of times as any other). The 'best' design has been

described as one with optimality (efficiency), level balance, utility balance, orthogonality and minimal overlap of levels of attributes. (Huber and Zwerina 1996). There is a trade-off between design efficiency (the ability of the design to derive attribute coefficients that are significant) and respondent efficiency (the ability of respondents to understand and complete the task) (Segal 1982). Recent work has suggested that the number of attributes and the number of levels can influence the results of the study (Ratcliffe 2002). However, Louviere (2001) reported that recent reviews of the literature outside health economics have been consistent in finding that task complexity and length primarily impact on the random component of variance not the mean parameters.

*"There is no empirical evidence that increasing numbers of attributes, numbers of choice options or numbers of choice sets (scenarios) impact mean preferences parameters, but there is evidence that increases in these factors impact random component variability. However optimal levels of these variables remain unknown for particular applications"* (Louviere 2001 pp34)

The importance of the random component in choice models is returned to further on in this chapter.

Also there is a "cost of thinking" (Shugan 1980). Respondents face different levels of motivation and have different levels of ability to respond to the choice task. One of the design questions to be explored is how to compare the complexity in decision-making in choice experiments with the complexity of other contingent valuation experimental methods. If it is similar, then other contingent valuation methods will face comparable problems. If there are additional levels of complexity in choice experiments then data and meaning may be lost by not understanding the nature or source of this complexity. However, it has been reported in health economics studies that simple models of choice have been found to explain 85% of the decisions made by individuals (Wordsworth, Ryan et al 2001, Ryan and Gerard 2001). However it has been argued outside the health economics literature that increasing the number of attributes would not significantly affect the estimates of the model coefficients of the model but that this increases the impact on the random component variability (Louviere 2001).

### **Assigning attribute levels to choice sets**

Assigning attribute levels is an area where there has been least guidance in the health economics literature to date, although outside the health economics literature the guidance

has been more comprehensive at dealing with these more complex issues of study design (Zwerina et al 1996, Louviere 2000).

The problem of maintaining the orthogonality of the design once the scenarios were put into pairs had not been addressed directly in the published health economics literature at the time of the study design. The technique of randomly assigning scenarios to pairs had been adopted in some studies (Ryan and Farrar 1995; Ratcliffe and Buxton 1999, Shackley, Slack et al 2000) or random pairing with additional manipulation to ensure trade-off between different dimensions of benefit (Farrar et al 2000). But subsequently there has been some suggestion that random pairing of levels of attributes may violate the statistical properties of the design (Ryan and Gerard 2001).

The other main approach has been to compare one scenario against the status quo (Ryan 2000, Farrar and Ryan 2000, Gosden et al 2000). This has the advantage of being cognitively easier and reflects the real world where one scenario is seen as the usual or current situation, practice or policy. A constant comparator can maintain statistical properties, but for this study where there is no constant comparator representing current practice or the status quo, this does not appear to be a logical option. Furthermore, by only comparing all levels of attributes with one attribute, the chance of comparing two scenarios other than with the baseline comparator, is lost. This important point (and how it was addressed in this study) it returned to further on in the next chapter.

### **Sample size**

The stated preference literature (in marketing) has tended to deal with the issues of sample size rather informally, with early empirical work suggesting samples of around 30 per subgroup were adequate (Cave et al 1993). The argument was that since these methods collect multiple observations per respondent, only small sample sizes were required (Green and Wind 1992). However, since the number of multiple observations provides more information on the individual and not the population, larger sample sizes are now more common. The figure of 75-100 respondents per market segment has been quoted (Bradley and Kroes 1990; Swanson et al 1993, unpublished working paper quoted in Cave et al 1993). The latter study tested how many respondents were needed to “uncover” known utility functions. It suggested that values could be predicted with the greatest accuracy as the sample approached 100, beyond which the marginal benefit of each additional interview falls.

The method of data collection also has an effect on the number of respondents who can be recruited to the study and complete the task. Postal questionnaires have been the dominant

method of data collection for other studies in the health economics literature. These have produced a higher number of respondents; between 100 and over 500 is the range (Ryan, Bate et al 2001), although one recent health economic study used data from only 67 respondents (Ryan 1998).

### **Tests of axioms**

Many choice experiment studies have included simple test of internal consistency by including scenarios considered to be dominant in all attributes, as reported in a recent systematic review of discrete choice based studies (Ryan and Gerard 2001). These additional choices are not “data” as the responses are not part of the orthogonal array and are not included in the regression model. There needs to be a balance between the amount of choice data from respondents that contributes to the analysis and the tests for rationality. If it is possible to present a large number of choice sets to individuals, or to split the sample into two separate questionnaires, as has been done in recent studies (Gerard 2002, Ratcliffe 1999), then the experimental choices and test of rationality should be included.

### **Data analysis**

Regression techniques are used to analyse responses to choice questions. These are data reducing techniques that attempt to predict an outcome based on as simple a model of independent variables as possible. It is a summary of the complex relationship between factors and is interpreted as an approximation of the true relationship that involves unknown variables. A good model is judged by whether it predicts a known outcome well (i.e. is unbiased) and predicts the outcome efficiently (balancing the number of explanatory variables which should be as few as possible, with the accuracy of the prediction).

Choice experiments provide a sample of repeated choice sets for individuals (choice between scenario A or B) for a given set of explanatory variables. The characteristics of the scenarios are the explanatory variables whose levels or presence/absence may contribute to determining a respondents’ decision about which scenario to choose. Interpreting the data from choice experiments in this way means it is possible to specify a regression function that describes the relationship between the attributes and the decision to choose a particular scenario. Because of the nature of the dependent variable, which is binary, simple linear regression does not function well for these kinds of data (Jones 2001, Powers and Xie 2000).

Responses to choice experiments can be interpreted as categorical binary dependent variables. Categorical variables are a limited rather than infinite number of possible values. If

the categorical variables are the independent variables (the attributes in this case) they can be assimilated into simple linear regression by using dummy variables to represent the presence ( $X=1$ ) or absence ( $X=0$ ) of the attribute. But if the dependent variable (the choice of scenario) is also categorical, then the analysis and interpretation of the data is more complicated. Since the dependent variable is either scenario A or scenario B, this is the case in choice experiments.

Classical linear regression models do not function well when the dependent variable is binary. This is because the predicted values of a regression function can lie outside the range 0 to 1. If the model is designed to predict a binary outcome, i.e. whether an individual might choose a scenario A over scenario B in, say, a pairwise choice, a value outside 0 (no) and 1 (yes) is theoretically impossible. Therefore other non-linear functions that are bounded to the range 0 and 1 are estimated. The common forms of models are the logit and probit models.

### **The probit model for analysing choice experiment data**

Probit and logit models are models that can estimate the value of dependent variables that are bounded within the range 0 to 1. They both have an S-shaped distribution and are similar in appearance. They differ in the assumptions made about the distribution of the error terms. As described above, probit models are based on standard normal distributions of the error term, assuming that the error term is the sum of independent unobserved quantities. The logit model assumes the error terms are independently and identically Gumbel distributed. Both are typically estimated by the method of maximum likelihood estimation that specifies the joint probability of observed data and determines the coefficient values that are most likely to fit the data.

The random effects probit model (Heckman and Willis 1976) has been a common option for analysis for health economists who have undertaken choice experiments since it is an approach that can take account of potential correlation between observations from one individual, but assumes that this correlation is unknowable, hence the random component. However, in some situations a simple probit model will be the most parsimonious model where this correlation is not significant. In practice, it is useful to test both ordinary probit, fixed effect and random effect probit models. The ordinary probit takes no account of any panel data (repeat data from the same respondent). The fixed effects probit model stratifies the data by respondent and assumes that each individual is different from others in a 'fixed' way. The random effects probit model assumes that the intra-respondent effects differ in a random way.



Using the random effects model that has been widely reported in the health economics literature, the function to be estimated in a choice experiment has the following form

$$U' = \text{constant} + \beta \chi_{1..n} + \theta$$

where  $U'$  is the (unobservable or latent) change in utility from moving from one scenario to another, and  $\beta$  represents all the coefficients of  $\chi$  attributes to be modelled.

The  $\beta$  parameters are equal to the marginal utilities of given attributes, and the ratio of any two parameters indicates the marginal rates of substitution between attributes, and  $\theta$  is the error term. The error term  $\theta$  is made up of a combination of unobserved heterogeneity ( $v$ ) and stochastic error ( $\epsilon$ ). The random effects probit takes account of any correlation between  $v$  and  $\epsilon$  which represents correlation between observations within any individual.

The approach assumes that there is a “latent variable” or underlying and unrevealed continuous variable  $Y$  that is represented by the binary variable 0 and 1. The coefficients of the independent variables relate to the underlying or linear index, often interpreted as the latent variable  $Y^*$ . This term is not directly observable and not measured by natural units, unlike the probability of choosing 0 or 1.

As mentioned earlier, the interpretation of the probit models is different from the linear probability models in that it applies a non-linear link transformation, using an S-curve of the normal distribution. The dependent variable  $Y$  can be interpreted as a *propensity* to take one action (in this case, to choose scenario A in a choice experiment, revealed as  $Y=1$ ). At a certain point along the continuous scale, the decision will switch to take the opposite action, (revealed as  $Y=0$ ).

One way to conceptualise this is in terms of costs and benefits relating to a decision. The ratio of costs and benefits in any decision is a continuous scale, but at a certain point, an individual will decide benefits are greater than costs (revealed as  $Y=1$ , the decision that A is a better choice than B). At a turning point, the costs of A will outweigh the benefits and the respondent will choose B (revealed as  $Y=0$ ). This is the latent variable approach to analysing categorical data. Therefore underlying each individual's choice is a continuous variable for the costs and benefits of each decision. All that can be observed is the decision to choose/not to choose in pairwise choice. This observable choice is a realisation of a continuous propensity that is unobservable, as formalised by random utility theory.

The estimated coefficients produced by the logit and probit model's coefficients cannot be compared directly as for the linear probability model. The two main findings in probit and logit models are the effect of the explanatory variables that is given by the sign of the coefficient. If it is negative, this means that the presence of this variable is less likely to make an individual choose a particular scenario in pairwise choice (in this case scenario A). This effect is shown by the size of the coefficient. The marginal rate of substitution, if appropriate, can be calculated from these coefficients.

The results of the random effects probit model can be estimated with and without a constant term. The constant term has been interpreted as the overall propensity for respondents to choose one of the scenarios even when the level of attributes is the same. If the constant term is significant, then respondents may be considering attributes other than those in the scenarios in their choice. A constant might not be specified in experiments where respondents are asked to assume that all other characteristics are the same except for those in the scenario, for example between two hospital outpatient clinics). The STATA (version 7.0) output from a random effects probit model automatically produces a regression equation with the constant term. The structure of the equation is:

$$U = \text{constant} + \beta_1 (\text{attribute 1}) + \beta_2 (\text{attribute 2}) + \beta_3 (\text{attribute 3}) \dots \beta_n (\text{attribute } n)$$

### **Interpreting the results of a choice experiment using the probit model**

The key findings of a choice experiment are the estimated coefficients for the attribute included in the design. These attributes show how important, in relative terms, each attribute is in an individual's (and aggregated) demand function. *A priori* reasoning provides hypotheses for the direction of the coefficients (whether they will positively or negatively affect the choice of a scenario). The p-value associated with the coefficient, and the confidence intervals provide information on whether a particular attribute is important in respondents' utility or demand for that scenario. The coefficients relate to one of the choices (scenario A or B) and a positive sign on the coefficient indicates whether increasing the level of an attribute makes it more likely that the scenario will be chosen.

The probit and logit models can also provide an estimate of marginal rates of substitution between attributes. Logit and probit models produce coefficients of a different magnitude but always in the same ratio, so the marginal rate of substitution is the same using either method (Powers and Xie 2002). If a price or charge is included as an attribute, the marginal rate of substitution between the price and another attribute may reflect the willingness-to-pay (to trade with money) for that attribute. If all attributes can be compared with one continuous

attribute, such as money (or travel time), this provides a marker for the relative and absolute importance of all attributes to an individual. This is subject to particular methodological constraints that are as discussed further on.

However, even without price as an attribute, it is still possible to estimate the relative importance of an attribute to the decision to choose either scenario A or scenario B. Therefore, while the probit model is complex, the results it can provide are relatively straightforward to interpret by non-experts. It has been argued that this is a potentially powerful tool for decision-making, and easily interpreted if analysed correctly (Jones 2001). However, the distribution of the error term, the use of a constant term, and the associated functional form of the estimated probit/logit model will lead to different estimates of the model coefficients.

There is no standard method of goodness of fit of probit models as there is for ordinary regression analysis. Since the coefficients are estimated using a maximum likelihood procedure that is not designed to maximise any one criterion of goodness of fit, there is no automatic method of assessing this (Powers and Xie 2000). The usual R-squared statistic for ordinary least squares regression (the proportion of the variance in the dependent variable which is explained by the variance in the independent variables) has no equivalent measure in probit models. The pseudo R-squared does not have the same relationship to probit models that the ordinary R-squared has to linear regression models. It is possible to artificially construct scenarios in which the pseudo-R-squared is very close to 1 even though the model is not a good fit. Conversely models with pseudo R-squared values that are very low can be very successful in terms of correctly predicting the observations in a sample (Greene 2003).

However, pseudo-R-squared statistics have been published for probit models. The most common is the McFadden's R-squared statistic (sometimes called the likelihood ratio index). This gives the proportional difference in log likelihood ratios of a model without parameters and a model with parameters, so that

$$\text{McFadden } R^2 = (\text{Log } L_0 - \text{Log } L_c) / \text{Log } L_0$$

where  $\text{Log } L_0$  is the log likelihood generated by the probit model with no parameters (i.e. with the constant term only) and  $\text{Log } L_c$  is the log likelihood generated for the current model. The statistic is presented as a percentage or proportion and indicates how well the model "fits" the data or predicts the dependent variable (in this case, whether scenario A or B is chosen)

(Altman 1991). The STATA computer package provides an estimate of the log likelihoods for probit models, so this statistic can be easily calculated from the data.

The McFadden  $R^2$  is a scaled measure that varies between 0 and (somewhat close to) 1. It is expected that the Pseudo  $R^2$  will be much less than what would be expected for a linear model. The Pseudo- $R^2$  in probit and logit models is best used to compare different specifications of the same model, rather than to compare models with different data sets. This approach to modelling is more open to error due to misspecification of the model and misinterpretation of the data than other simpler regression methods. There is no standard accepted range for the test of goodness of fit. This issue is referred to again in chapter 7 in considering the goodness of fit of the model. Since there is no widely accepted  $R^2$  statistic for binomial data, the convention is that these measures should be used “cautiously” (Altman 1999).

The approach adopted for this analysis will be to use this statistic to assess which of the models presented is the better fit to the data *relative* to one another rather than to interpret the statistic as a measure of the overall goodness of fit of the model.

### Validity

The purpose of hold-outs is to determine the model's ability to correctly predict the 0/1 observations in the sample in a set of pairwise choice as this may be a more reliable and useful test of the model than the pseudo-R-square statistic. The hold-out approach is a way of testing the internal validity or robustness of the model. The hold-out choice sets have the same structure (ten choices, five attributes each) as those included in the main study derived from an orthogonal array of choice sets. Respondents are presented with additional choice sets but the responses to these choice sets are not used to estimate the regression model. The test is whether the model will correctly predict the actual choices that are made by respondents. Once the model has been specified using data from the main experiment, the choices made by respondents to the hold-out can be compared with the choices predicted by the model, using for example a chi-square test. If the model is well specified, then the actual proportion of patients who chose a scenario would be similar to the proportion predicted by the model. The chi-square test indicates the proportion of agreement between observed and expected values.

Tests of external validity (the test of whether respondents' true preferences are found using a choice experiment approach) are more difficult in this context. Louviere (2001) has pointed out that the efficiency of any choice experiment depends on knowing the true parameters,

but if these were known *ex ante*, then it would not be necessary to run the choice experiment regression model. For experiments where a willingness-to-pay estimate is derived, external validity could compare the values derived from the choice experiment with real market values where they existed. However, Louviere (2001) suggests that the estimates of value derived from choice models have often been over-simplistic due to a lack of understanding of the true nature of the error term in the models. Louviere suggests that the random error term incorporated into most choice models may be derived from different sources of error. He argues that reasoning and empirical research imply that the random component is multidimensional, containing subcomponents representing intra- and inter- individual variation, task variation and potentially many other sources specific to a research context.

Louviere has argued that the variation in values reported by different studies may be due more to the differences in the error term, than the differences in the actual values. This violates the condition that the variation in error terms between two utility functions must be less than the variation in parameters of the explanatory variables. For monetary values, this point is illustrated by the reported differences in willingness to pay (WTP) and willingness to accept compensation (WTA) values. Louviere has suggested that the random components may be different between WTP and WTA since WTA values have been found to be consistently larger than WTP values,. This may have more to do with the fact that the variance is wider for WTA than for WTP since subjects are less familiar with the concept of deciding a level of financial compensation for forsaking something they value, than the maximum price they would pay it.

Therefore much remains unknown or unresolved, particularly regarding optimally efficient designs for particular purposes and the degree of complexity that is desirable and/or necessary to understand adequately, explain and predict behaviour.

### **When should choice experiment methods be considered in evaluation?**

#### *To consider more than one trade off*

Many of the issues that arise in trying to use CE to obtain monetary valuations of outcomes also hold for a wider class of contingent valuation techniques. Clearly, the importance of the arguments about whether and how to include monetary valuation will depend on the health care setting and the evaluation question. Also it is determined by whether and how it can be included in a way that makes sense to the target group of respondents who are asked to make the choices in the choice experiment setting.

It has been argued that choice experiments represents “*a subtler way to establish WTP, and hence the monetary value of benefits*” (Ryan 1996). One recent study compared the choice experiment approach with that of a study examining patient preferences for the same treatment using a standard gamble technique (Morgan, Shackley et al 2000). They found that while the standard gamble (SG) findings supported those of the choice experiment (strong preferences for local treatment), there were differences in the level of risks that patients were willing to take to ensure local treatment. The standard gamble study found that patients were willing to accept higher risks of dying than in the choice experiment study. There were notable differences in definitions presented to patients in the two study designs making this comparison less valid. However the authors suggest that the SG approach presented the trade-off in more explicit terms than the CE approach as the SG approach trades off only two attributes at a time. By contrast, in the CE approach these attributes are just two among several other attributes for the respondent to consider. Therefore, the choice of experimental design must depend on the nature of the question and how realistic the trade-off is that respondents are required to make.

*To mirror real management problems in the health service*

For some health services research questions, the consideration of multiples of attributes at the same time is seen as an improvement over methods that consider only one dominant outcome. The argument is that choice experiments can ask respondents to consider a wider range of consequences than other trade off methods (Ryan 2000). For areas of health policy that need to consider the process of care or a range of dimensions of outcome, this could be a valuable property of CE methods. The services where CE has been applied recently have focused on patients' preferences for aspects of care other than a single health outcome. Examples of this were a study of miscarriage management (Ryan and Hughes 1997) and a study on patients' preferences for out-of-hours care (Morgan, Shackley et al 2000). These methods have also been used to investigate the importance of waiting time compared with other aspects of care (van der Pol and Cairns 1998, Ryan, McIntosh 2000), to explore job satisfaction in general practice (Scott 2001, Gosden, Bowler et al 2000), and preferences for different types of investigative procedures (Bryan et al 1998).

Studies that do not include a price or “charge” for health care have been published recently (Ryan, McIntosh 1998, Morgan, Shackley et al 2000, Shackley, Slack et al 2001, Longworth, Ratcliffe et al 2001, Moayyedi, Wardman et al 2002). These studies have focused on patients' preferences for the organisation of health care where the relative value of specific attributes is not known and is considered to be important for management decisions, for example the value of introducing a patient health card in primary care (Ryan et al 1998).

Where resources are used across a range of different activities and little is known of the relative value of these different aspects of service to users (or the general public) these methods may be especially useful. One study used two choice experiment designs, one with a price attribute, one without (Bryan et al 1998). The argument given was that there might be some objection by respondents to considering a charge, and asking respondents to consider a price for health care would make the experiment unrealistic. When included, the price attribute was not found to be statistically significant suggesting that patients did not consider the price in their decision-making.

### **Choice experiments and willingness-to-pay**

One of the potentially most interesting uses of CE is the possibility that monetary valuation of the value of individual attributes might be derived from the technique (Ryan, Bate et al 2001). If one of the attributes used in the experiments is price or a charge (some notion of financial sacrifice), then this allows willingness-to-pay to be indirectly derived from the marginal rate of substitution between an attribute and price. This approach has been adopted as the preferred approach to estimating WTP values by economists working in other public sectors, rather than to ask them to state their WTP directly (ibid.)

However, this method of estimating WTP values does not avoid other problems contingent valuation methods in that individuals are often not accustomed to valuing outcomes or services in terms of money and that ability to pay biases stated WTP. One study that stratified respondents receiving treatment for infertility into income groups found that respondents in the higher income group had a lower marginal valuation of price, reflecting diminishing marginal utility of income (Ryan 1999).

There has been some controversy over whether differences in WTP are affected by choice of payment vehicle. Cost-benefit theory would suggest that payment mode is a preference and therefore should be considered if all elements of a policy are to be valued by respondents (Boardman et al 2001). Empirical studies that have investigated the use of a price attribute and type of cost vehicle have recently been undertaken have not been conclusive. In an Australian study, the cost attribute was described as a Medicare levy (Jan et al 2000). The arguments given for using this method of payment were that respondents would have little experience for paying for hospital services. The Medicare levy attribute was not found to be statistically significant. One of the reasons for this suggested by the authors was that since over half the respondents were not in paid work, many would not be paying such a levy in real life. They assumed from this that respondents might have discounted this

attribute altogether in their choice-making. In this case, the payment vehicle may have had an impact on whether cost was considered by individuals.

A recent systematic review of current practice reported only 55% of the studies included a monetary attribute from which a marginal WTP could be calculated (Ryan and Gerard 2001). Issues were raised about the appropriateness of price as a proxy for monetary valuation, especially in a collectively funded health system (as in the UK). There appears to be a trend away from directly applying WTP attributes in discrete choice studies, but attempts to use other indirect monetary valuation (Jan et al 2000) have shown that respondents were not sensitive to these values in their decision-making.

Another more recent Danish study has found disutility associated with both payment as a concept and the extent of payment through different payment vehicles and that these disutilities are separable. (Skoldborg and Gyrd-Hansen 2002). Respondents were more positively inclined when presented with the concept of introducing out-of-pocket payments than towards the notion of a tax increase. However the marginal disutility associated with a rise in out of pocket payments was greater than that for taxes.

Another problem is that choice experiments, rather than open-ended WTP methods, may not derive individuals' maximum willingness to pay (i.e. their threshold of indifference) and may underestimate it (Ratcliffe 2000). This problem was also found in the Danish study (Skoldborg and Gyrd-Hansen 2002). This study found that some respondents would still choose a particular scenario even when associated with very high costs (which was only included for a sub-group of the respondents). It concluded that if very high estimates of value are used, respondents will still choose these scenarios and the WTP will be estimated far higher than if only lower estimates were used.

This evidence suggests that patients do not seem to consider the price attribute in their choices (Bryan, Buxton et al 1998, Jan, Mooney et al 2000). Using proxies for the charge for care, such as travel cost to individuals (Ryan, McIntosh et al 2000) appears to have been more successful as these costs may actually have to be borne by patients using the service and therefore make more sense in a decision-making exercise. Similarly, the income that GPs would be prepared to sacrifice in order to improve their working conditions (Gosden, Bowler 2000) makes sense to respondents who face these financial decisions in the real world.



### **Willingness -to-accept attributes in choice experiments**

The OPUS study referred to in the previous chapter developed an outcome measure of social care for older people, similar to the EQ-5D but measured in different domains of quality of life (Netten et al 2002). The study adopted a choice experiment approach to identify weighting for particular domains of care that were important to this target group. The appropriateness of including a monetary charge as an attribute was considered in the choice experiment in order to establish a WTP for each level and domain. The authors argued that this would have the added benefit of showing that the weighed measures in the outcome instrument were cardinal. They assessed the feasibility of ascertaining appropriate levels for the monetary charge but this did not work well. Open-ended valuation in the pilot stage was *“virtually impossible and people did not like the very wide differences between levels or very high levels of possible payments being presented”* (ibid. pp29).

Another barrier cited for this method was ability to pay, as this quote demonstrates:

*“The initial interview with a woman who was on a very low income made it clear that she could not afford anything other than her current living expenses. This raised the concern that, if asked to make choices including levels of payment at anything other than very low levels, the money issue would dominate, so whatever option was presented, the cheaper option would always be chosen.”* (Netten et al 2002 pp30)

This is an important finding and highly relevant to the decision of whether to include a charge in the palliative day care study. Even though this study was published after the design of the PDC study, it reflects the same concerns about adding monetary considerations to decision-making by people who are mostly elderly. The approach finally adopted for the OPUS study was to use a willingness-to-accept approach to establish the financial recompense or benefits that would make an individual indifferent between living in a more disadvantaged situation with the financial benefits, and the more advantageous care situation.

## **SECTION 6. THE RELEVANCE FOR THE EVALUATION OF COMPLEX SERVICES**

The choice experiment method has a number of advantages for examining the value of complex health and social care services. Choice modelling can be used to estimate the structure of preferences for these services, that is, what is contributing most and least to the decision to access care. This re-directs the question away from one that is very difficult to answer, like “is a service cost-effective overall?” towards a more answerable question that is valuable to decision-makers.

An analysis of the demand for complex areas of health and social care using choice experiments may reveal the characteristics of the service that are important to patients. In a service where external measurements of health gain may not be as important as the subjective assessment of quality of life (welfare/utility), this approach to measuring outcomes can be a way forward. It can therefore make progress in understanding the nature and value of the outcomes of a multidimensional service in ways that are not amenable to other economic evaluation methodologies.

The choice experiment methodology allows the question of effectiveness to be approached from a different perspective. Rather than measuring the impact of a service on an individual's health as a proxy for utility, choice experiments can evaluate the services directly by establishing how much individuals would be willing to sacrifice in order to have these services.

But there is also another potential use of this approach that is of specific interest in the context of PDC and is probably equally as important in other complex services as well. There is debate in the PDC field about the relative importance of the various activities it offers. CE techniques can be used to elicit patients' preferences for particular component activities of care that are difficult to obtain by other methods. They can by-pass the objective measurement of health status and consider individuals' preferences for particular characteristics of a service directly. Measurement of preference for inputs or outcomes amounts to measuring the same change in welfare. For example, the attributes of the service (or arguments in patients' demand function for PDC) may be described either as the objective characteristic of the service (information provided, presence of a doctor) or as the subjective perception of benefits from a specific attribute of a service (better clinical information or less time in pain).

For PDC research, this reasoning addresses a fundamental problem: that objective measurement of outcome may not be the appropriate way to judge the value of the service to society. If, using choice experiment methods, the value (or lack of value) of PDC services can be established, then this will be a useful contribution to knowledge and decision-making. If the choice experiment included a monetary valuation of PDC as an attribute, the results could be used in a cost-benefit analysis of PDC.

### **Adopting a choice experiment to palliative day care**

The pilot study for the North Thames study indicated that patients are willing and able to make choices about the service that contribute to their welfare (Douglas et al 2000). PDC

patients were encouraged to make choices about the kind of care they accessed and had a range of activities and therapies offered to them. Second, staff involved in that study reported that patients demanded particular services (such as hairdressing) and that they might be willing to pay for them if they were not provided free of charge. Third, although the majority of patients were over 65 years and had advanced illness, they did not suffer from cognitive impairment as this is often one of the criteria for access to PDC. Consequently the study showed that such patients were usually very willing and able to answer questions directly and are used to being asked their opinions.

Since PDC is not easy to describe in a straightforward way, there is little clarity about what the attributes are from a patients' perspective. This means there are problems with translating PDC into specific attributes of the service. Some of the characteristics such as "friendly social environment" are, in reality, a bundle of different attributes, and can mean different things to different people. Furthermore, finding out that patients want a "friendly social environment" does not translate into policy advice that is straightforward or easy to action. Providers need information that is specific so that the important attributes of the service that can be defined (and purchased) by them. This means that the focus of the study becomes provider driven rather than user driven. There is a danger, then, that the value that patients derive from attending PDC may not equate with service provision type characteristics and may include other factors. Attributes such as "staff attitude", "friendliness", "and environment" may have more importance than any of the characteristics presented in the choice experiment. This could be seen as a limitation of this approach and is raised again later in the thesis.

### **Incorporating a price or charge as an attribute**

Earlier in this chapter, the arguments for and against a willingness-to-pay approach were set out. The problems were that the valuation may be influenced by respondents' ability to pay; they may not be able to express their preference in monetary terms; there may be political dissent from service providers and the public in valuing health services in this way and the approach may not derive maximum willingness to pay (i.e. the point of indifference in a trade-off) if the price attribute is set to levels that do not reach this maximum. The willingness-to-accept approach adopted by the OPUS study would have to ask respondents to consider financial recompense for worse quality of life before death.

In the context of PDC, all these problems were thought to be important, so much so that it was felt that the inclusion of a price attribute in the first study of its kind in PDC could be so controversial as to potentially jeopardise the whole project. PDC centres do not charge for

use of their services (except a nominal fee for lunch in some centres). Many of the services (general and specialist) are provided by volunteers. All but a very few of the people who attend these centres are not in paid work and are predominantly elderly (Higginson et al 2000, Goodwin et al 2002). The concern was that, in this setting, the inclusion of a charge for attending PDC could provoke so much protest by providers of care to the extent that they would not support the study, especially if it was seen as a way of setting charges for PDC “by the back door”. This could be seen to be against the ethos of the service.

Second, although no data had been systematically collected on patients’ income in the previous studies, it might be the case (as in the OPUS study) that the consideration of a charge for PDC would be dominant in many patients’ consideration of their choice, and that data on their value of other aspects of the service would not be obtained. The decision taken was that a price attribute would not be included in choice experiment since there was a danger that the experiment could fail overall. Given the difficulty that had been experienced in previous studies in obtaining *any* evidence of the value (or lack of value) of PDC, the aim was to obtain some data on the relative value to patients of specific aspects of PDC rather than risk obtaining none at all.

## **SECTION 7. AIM OF THE EMPIRICAL RESEARCH – HIERARCHY OF OBJECTIVES**

Since it had not been investigated empirically, it was decided to test the hypothesis that EQ-5D was not sufficiently sensitive to detect any changes in health status between patients who attended PDC over a three-month period compared with patients who did not attend. The EQ-5D instrument is a relatively easy questionnaire to administer to patients, with only five items and a visual analogue scale, so it was not thought to represent a considerable burden to respondents. However, the long-term goal of the research was to be able to compare PDC with other services for patients in a palliative phase of illness (home care, social care, primary care, and inpatient care) and to establish the marginal costs and benefits of the service. However in order to reach this goal, or to make progress towards it, other stages of research needed to be undertaken.

Choice experiments offer the opportunity to investigate the (putatively) important constituent parts of PDC and to quantify the relative importance of them to individuals’ utility or overall welfare. This seemed to be a sensible approach to evaluate an intervention that aims to meet a range of needs. The application of this method in a PDC population had not been tried out before. In light of this, and in the context where no other method had shown any useful quantifiable benefits of PDC, it was appropriate to design an experiment that had the highest chance of producing useful data and minimise the risks of obtaining no data at all.

A hierarchy of objectives for the evaluation was drawn up. The first was to establish whether a stated preference method could be used in the context of PDC. It was not clear beforehand that it would be possible to use this technique with PDC patients. Therefore, it was necessary to demonstrate that the techniques could arrive at some relative measure of value for attributes of PDC, and to address the ethical and practical problems of undertaking research on frail and elderly people who use PDC services.

Second, the attributes had to be chosen that would be useful to policy-makers. This meant identifying aspects of palliative care where there was little evidence of value and that were not the same in all centres. From this, we could establish what seemed to be more valuable to users and whether particular characteristics of the respondents or the service had an impact on their valuation of specific attributes of PDC. For a service that is not uniform and where the provision of care changes relatively rapidly, this could be an appropriate way of assessing the most preferred package of services (or for individuals with particular characteristics). The findings would be of value to providers where this information has not been obtained through other methods of evaluation.

Only if the first two objectives could be met would it be useful to consider whether a monetary valuation of PDC attributes could be derived using this method. The focus of the empirical work was on the first stage and second stage of investigation. The question of whether it would be possible to reach the third stage in this study and quantify the value of PDC attributes in monetary terms was considered and rejected at this stage as it was thought to be not feasible to identify an appropriate measure of sacrifice for this group of patients in this context. This decision will be discussed in more detail in the next chapter and in the conclusion. The fourth wider objective was to consider whether the empirical analysis would provide any insights for the economic evaluation methods of health and social care interventions with similar characteristics (multidimensionality, and complexity of outcomes).

The choice experiment study was therefore designed to measure the strength of preference for different attributes of PDC as a measure of *relative* rather than absolute benefit of PDC. This approach measures the marginal value of each attributes relative to all other attributes. This is still based on the theoretical underpinnings of CBA (Mishan 1988; McIntosh et al. 1999) but is one step removed from valuing all attributes with reference to the 'measuring rod of money'. It is a step along this path that has not been undertaken before in attempting to quantify the value of PDC services.

## SUMMARY

This chapter has reviewed the health gain and preference-based approaches to economic evaluation, with a particular focus on the EQ-5D method of measurement and on choice experiments. These methods were discussed in depth and some of the possible methodological problems identified. The main problem with the EQ-5D instrument is that it may not be sensitive enough to capture small but subjectively important changes in health status. In the context of palliative care, the value of time at the end of life may be qualitatively different from the value of time earlier on in life, which argues against health gain being measured in additively separable and equal units of outcome.

The choice experiment approach has also been considered as a method of assessing the relative value of particular dimensions of care. In the context of multidimensional services this is a promising approach since the focus on the strength of preference for particular aspects of care rather than on overall quality of life that is hard to capture using simpler measurement instruments. Stated preference choice experiment methods can adhere more closely to the theoretical underpinnings of CBA (welfare economics) than health-related quality of life assessment.

Choice experiments are considered to be worth investigating in complex services as they have other properties (such as simplicity of the task, and data collection at one time point only) that are important for collecting data from groups that may be frail, confused or whose health is deteriorating. The next chapter outlines the research methods adopted in the empirical investigation of these issues presented later in the thesis.

## Chapter 5

**Methods of empirical research into palliative day care****Introduction**

This chapter describes the methods of the economic evaluation of palliative day care in a comparative study undertaken in Chichester and the south of England. It describes the two components of the economic evaluation that form the empirical research presented in this thesis. This follow-on study was different from the first North Thames Palliative Day Care study in a number of dimensions that will be described here.

The North Thames palliative day care study had included a detailed cost study to compare the range of inputs (day care and other health and special care services) and cost of care between patients who attended and those who did not. The same methodology was adopted for the Chichester study.

The two distinct parts of the methods of empirical investigation are described separately, as they were in the previous chapter. These two stages of research were undertaken consecutively and the same patients did not participate in both stages of data collection. As in the previous chapter, the choice experiment methods are described in more detail since these are less well known and, at the time of planning the study, there were differences in the reporting of methods of choice experiments in the health economics literature. The section on choice experiments refers to the literature that was published before and during the period of planning the study design, and discusses the ways in which this study follows or diverges from the methods used in other health economic studies.

**Background to the Chichester study**

This study of palliative day care (referred to from here on as the Chichester Study) differed from the previous North Thames Palliative Day Care Study in that the quality of life data was gathered from a prospective before-and-after study of one palliative day centre in the south of England where palliative day care provision did not previously exist. An opportunity arose to recruit patients before a new palliative day care centre was built (attached to an inpatient unit) and then recruit a palliative day care group after the centre opened. A third group was also recruited at the same time to assess any changes in local provision of service (such as home care nursing services) between the 'before' and 'after' PDC groups.

This study evaluated the impact of a new PDC service in a locality where one had not existed before. This was a more robust study design than the North Thames study that had recruited patients who attended five PDC centres and the comparison group was patients who did not wish to attend PDC in the same localities. This approach was more open to problems of sample bias than the before-and-after study as reported in the literature review in chapter 3.

A preference-based health-related quality of life instrument was added to the set of disease-specific quality of life questionnaires that had been used in the North Thames study. The choice of questionnaire was determined from a review of existing outcome instruments designed for economic evaluation which concluded that

*“the EQ-5D and HUI [Health Utility Index] are currently the best preference-based health status measures and should be considered for inclusion in all trials intended to be used in economic evaluation.”* (Brazier and Deverill 1999, pp 4)

The HUI family of health status measures has been described as incorporating “within the skin” definitions of health status, which is a person’s capacity rather than actual experience of living in a health state (Furlong, Feeny et al 2001). EQ-5D incorporates the consequences of ill-health (ability to undertake usual activities, social functioning). For this reason, it was considered that the EQ-5D might be the more appropriate measure in principle the context of palliative day care.

The health-related quality of life questionnaires were also included in the Chichester study to assess whether the more robust study design would be able to detect differences in outcome using these outcome instruments. This was undertaken in the Chichester PDC centre only. The choice modelling came after the quality of life study. It did not focus on one PDC centre only. Patients from four PDC centres were recruited as not enough patients were attending the Chichester PDC centre to gather sufficient data. This also meant that the study could investigate whether patients at different models of service had different preferences for models of day care (i.e. whether provision determines preference) and to do this, patients were recruited from centres offering different ranges of services.

As the Chichester study was a collaborative project, it is necessary to set out the specific contribution of the author of the thesis (HRD) to the empirical research. The design of the cost study and the collection of all the cost data were undertaken by HRD. The published version of the EQ-5D questionnaire was used and the structure and design of the



questionnaire was not changed from the published version. The choice experiment was conceived and designed by HRD with additional statistical advice from a colleague in the Health Services Research Unit, LSHTM.

## **Section 1. Overview of the Chichester study**

### **Methods**

A prospective before-and-after study was undertaken to compare the costs and outcomes of three groups of patients and their carers. The before group was comprised of all patients receiving a palliative care service (inpatient care or specialist home care) before the day centre opened. They were receiving usual palliative care services (in-patient, out-patient and home care) and would be suitable for day care, as decided in weekly hospice team meetings. These patients lived at home and were selected as having specific needs for psychological, social, nursing or physical care, or support for their carer that could be provided in a day care setting, although one did not yet exist. The day care group consisted of all new referrals to palliative day care once the service commenced. The comparison group was defined as those patients who did not receive day care when it opened, either because they were not referred or did not wish to receive it, but they continued to receive usual palliative care services.

### **Setting**

Chichester, West Sussex is a district within the South East Region. The district has a resident population of 759,000 of whom 156,000 (20.1%) are over 65 years of age. The hospice has 15 inpatient beds, plus a home care service caring for patients and their families in the community. Around four hundred patients were admitted to the hospice each year, and in addition there were 550 doctor visits and approximately 4,000 home care nurse visits to patients in their own home per year, during the study period. Half of the referrals were from hospital doctors and half from general practitioners.

### **The day care service/ intervention**

During 1999, the hospice had been planning the development of a purpose built day centre. This opened in February 2000, for one day per week, taking approximately 10 patients. This expanded to three days per week, taking 12-15 patients each day from home care and the hospice in-patient unit. The service included social activities, plus nursing & medical care, physiotherapy, occupational therapy and volunteer support. Patients usually attended one day per week, although some who needed more intensive support attended two or three times per week.

## **Data collection**

Data were collected using trained interviewers. Interviews were undertaken at baseline which was either entry to PDC or for the before group at the point they agreed to take part in the study; at 6-8 weeks, and at 12-15 weeks. The interview took place in the patients' homes, by arrangement. At each interview, the EQ-5D questionnaire was administered. Two other disease-specific quality of life instruments were also used. The previous multicentre study informed the choice of outcome measures that were selected to evaluate palliative day care. The Palliative care Outcome Scale (Higginson and Hearn 2000) was used to allow comparisons with the previous study. However, the McGill Quality of Life Questionnaire (Cohen et al 1997) was not used in this study since it was longer (16 items versus 10 in POS), there was more missing data and some patients found the items distressing (Goodwin et al 2002b). Therefore a measure was selected that reflected the existential objective of day care, engendering hope. The Herth Hope Index (Herth 2001) was piloted for the first time in palliative day care in this study. The results of these palliative care quality of life measures are not reported here since they were analysed by others and did not form part of the economic evaluation.

## **Sample size**

The sample size for the study was based on 5% level of significance (two-tailed t-test). A sample size of 40 in both intervention and before or comparison groups would give a power of over 85% for the disease-specific measures. The power needed to detect difference in EQ-5D scores was not considered. The power calculation for economic analyses has been shown to be different from that of clinical studies since the associated variance and covariance is not usually the same when costs and effectiveness are synthesised in cost-effectiveness ratios (Briggs et al 2002). This was a limitation of the study design (as discussed further on in chapter 7).

To ensure a chance of significant findings in the clinical effectiveness study the aim was to recruit 70 in each of the three groups – 210 patients in total. However, considering the number of new referrals to day care (approximately 5-10 per month) it was not possible to recruit 70 patients to the day care group. The sample size was recalculated at 170 patients.

## **Eligibility criteria**

All patients were asked to be in the study except those in the following group: patients too ill to be interviewed (decision to be made by the clinical team), with confusion/ dementia; with behavioural problems which would exclude the patient from day care; those currently out of the area (e.g. staying with relatives elsewhere; those reluctant to have further contact with

the hospice; those under eighteen years of age; patients without malignant disease or other form of chronic life threatening illness.

## **SECTION 2. Methods of estimating resource use and costs**

Analysis of costs of care between the three palliative care groups (palliative day care group, 'before' group, 'contemporary comparison group') was undertaken to compare the overall level of costs over the period of recruitment to the study, whether they changed over time within groups, and whether there were significant differences between groups in mean and standard deviation of costs.

Data was gathered on the full range of services used by the patient and family using a resource use interview and additional data on use of hospice care extracted from clinical records. From this information the total volume of cost generating events was calculated. Data on inpatient and outpatient services, residential care, use of primary and social care, services to support families, were collected. Using a vector of unit costs estimated in the North Thames study (for PDC and for hospice care, see below) and from published data for all other health and social care services (Netten et al 2002), the total cost of care during four weeks prior to interview was calculated for each patient. This allowed comparison of costs over time in each of these resources use "windows" for each arms of the trial, before day care, the palliative day care group, and the contemporary comparison group.

Prior to the North Thames study, there had been no detailed analysis of the resources dedicated to PDC and costs. A 'bottom-up and top-down' method of costing was employed in order to consider the range and volume of resources used in the centres, rather than rely solely on financial data. Detailed cost data were collected from all five centres in the North Thames study and the same unit costs, uplifted to 2002 prices were used for the follow-on Chichester study.

A second level of analysis was undertaken to assess whether particular patterns of resource use were responsible for differences in costs between groups. Two individual health care resources were assessed: inpatient use was analysed as a separate cost item (as it is expensive and therefore an important cost driver for patients who use palliative care services). Also, the use of GP services was analysed to assess whether palliative day care acted as a substitute for patients' access to GP services. It was hypothesised that patients who use palliative day care services might use less GP services if they were receiving specialist support and symptom review at the palliative day care unit.

### **SECTION 3. ASSESSMENT OF OUTCOME USING THE EQ-5D INSTRUMENT IN THE CHICHESTER STUDY**

The properties of the EQ-5D instrument described in the previous chapter suggest that it might not be sensitive enough (with only 5 domains and three levels for each domain) to identify some of the aspects of quality of life that are important in palliative day care, such as engendering hope, 'existential' health, and rehabilitation. As there has been no empirical study undertaken in this health care setting to explore these issues, the EQ-5D instrument was adopted for the Chichester study.

The aim of the study was to establish whether there was any difference in quality of life between patients who attended the PDC centre in Chichester compared to historical controls, recruited in the months before the centre opened. The null hypothesis was that EQ-5D would not detect significant differences in quality of life.

#### **Analysis plan for the EQ-5D data**

The first stage was to report the numbers of respondents who reported no problems, some problems, or extreme problems in each health domain. This was set alongside the same scores for the population of the UK to show how these palliative day care patients differed from this profile of scores.

The quality of life weights were calculated for all palliative care groups at each stage of interview. These overall scores were derived from the scores on individual domains. Values for the 243 possible health states defined by the EuroQoL classification have been calculated using a regression model developed by the EQ-5D group (Kind et al 1998). The result of this arithmetic model is that it allows the scores for the five domains to be translated into a composite number, as described in chapter 3.

The composite quality of life scores (EQ-5D score) derived from this arithmetic model were examined to assess whether any significant change in score over time could be detected. The median, quartiles, confidence intervals and extreme values are reported in box plots that show the different patterns of data in each palliative care group at each time point. Within each group, respondents who had data for interviews one, two and three were examined separately from patients who only had data for interviews one or interviews one and two only. It was hypothesised that these groups represented patients at different stages of disease (as defined as time from death or acute decline. Changes over time were compared between interview one and two; and between two and three.

Paired t-tests are reported for each of the palliative care groups. The t-test examines the differences in mean values of health status for each respondent at each interview and test whether the average scores is different from zero.

The visual analogue scale (VAS) scores are reported for completeness but these are not included in the derivation of the QALY weightings. This measure is often omitted from clinical trials as it does not contribute to the overall outcome score. However, it represents a descriptive snapshot of time for patients in the palliative phase of illness that has not been reported in this way before for palliative day care, so it has been included in the results. It may also help to interpret and contextualise the overall quality of life score if they do not show a clearly interpretable pattern. As reported in chapter 4, VAS scores have been interpreted as representing the 'morale' of respondents at a moment in time, that is, how they are feeling overall about their quality of life, not specifically related to their health, but in the wider context of their lives (Williams 2000). This may, hypothetically, be more closely related to the objectives of PDC than the five domains of health-related quality of life that contribute to the overall score.

The main comparison in this study is between the palliative day care group and the group of patients who were recruited before the palliative day care centre was opened (the "before" group). If there is a significant difference detected between palliative care groups, it could be argued that this is related to the different times in which patients are recruited to the study. The overall health economy may have altered (for example changes in community or primary care services) resulting in type one error in the analysis (false positive associations between intervention and outcome). To address this, the following analysis plan was adopted: if a difference in EQ-5D scores were to be detected, then a second level of analysis would be undertaken with the contemporary comparison group (patients recruited at the same time as the day care group but not attending day care). This comparison would only be undertaken if a difference between the before and after group were detected.

### **Validation of the EQ-5D scores against data disease-specific health-related quality of life instruments**

At the same time as collecting data for the EQ-5D instrument, respondents were asked to complete questionnaires relating specifically to palliative care (the Palliative Care Outcome Scale and the Herth Hope Index). The Herth Hope instrument provides statements to which respondents must strongly agree/agree or strongly disagree/ disagree (four options). These statements describe feelings about the world around, and attitudes to life and to the future. None of the domains coincided with the domains in EQ-5D. On the other hand, the POS

instrument has been designed as a clinical tool. It includes questions on physical symptoms, psychological symptoms, spiritual considerations, practical concerns, emotional concerns and psychosocial needs. Two questions in POS relate closely to questions in the EQ-5D instrument. The anxiety/depression statement in EQ-5D ('I am not/moderately/extremely anxious or depressed') is close to a POS question ('over the past 3 days have you been feeling anxious or worried about your illness or treatment?' - response choice: not at all, slightly, moderately, severely, and overwhelmingly). Similarly, the EQ-5D pain question ('I have no/moderate/extreme pain or discomfort') is similar to a POS question ('Over the past three days have you been affected by pain?' - same response choices as before).

The data on the responses to these questions answered by the same group of respondents at the same interview were compared in SPSS (version 10) using the Pearson correlation coefficient. This expresses the degree of linear relationship between two variables measured from the same individual. Values can range between -1.00 to +1.00. A correlation coefficient of +1.00 signifies a perfect positive relationship, while -1.00 shows a perfect negative relationship. The smallest correlation is zero. Significance at the 1% and 5% levels were explored.

### **Qualitative data**

In-depth qualitative interviews were not conducted for this study, but patients were asked three questions at the end of the interviews: what is day care like, what is the most important thing about day care, and whether they felt they had changed since attending day care. The responses were analysed by another researcher on the study team. The responses to the question 'have you changed as a result of attending palliative day care?' are included in the results section as they provide some insights into patients expressed preferences. The contrast between this approach and the constrained choice approach of choice experiments will be discussed in the final chapters.

### **Informed consent and ethics**

Local ethics committee approval was obtained before data collection began. The application for consent was undertaken by Professor Irene Higginson, principal investigator for the project at Guy's Kings and St Thomas' Medical School. One of the particular ethical issues was the need to consider the inclusion of all patients who might attend a palliative day care centre. These include patients with motor neurone disease who were not able to provide written consent. In these cases, written consent was given formally by a carer and signed witnessed by a third party (see Appendix B for copies of the consent form). No patients were excluded from the study on the basis of their disability. The informed consent letters

were held in the Department of Palliative Care at Guy's Kings and St Thomas' Medical School.

The aims and objectives of the study were discussed with patients and an information sheet about the study was left with them so they could consider whether they wanted to take part. The interviewers were nurses who worked part-time in the hospice. Since none of the questions related directly to care outside PDC and since none of the nurses worked in the PDC unit, there were perceived to be no potential conflicts of interest between patients' quality of care and taking part in the study. On the contrary, the fact that the patients knew that the interviewers worked for the hospice (although they did not care for them directly) was thought to have contributed positively to patients' decisions to take part in the study. The recruitment rate was higher than that achieved in the multicentre study.

### **SECTION 3. THE CHOICE EXPERIMENT**

This section describes the methods and practical considerations associated with undertaking the choice experiment in PDC. For each stage in study design, the method adopted in the PDC study is presented, alongside evidence from other choice experiment studies in the health economics literature to demonstrate that the methods are following an established methodology for choice experiments, where this exists. In some areas (such as establishing design balance), the PDC study took a different approach from other studies and this is explained and justified.

#### **Aims and hypotheses to be tested in the PDC choice experiment**

The aims of the study were to establish that a choice experiment study could be carried out in a PDC setting, to estimate the relative importance of attributes (services and structure of care) of PDC, and to evaluate whether demand for particular attributes of the service is influenced by patient characteristics or particular models of care. This was evaluated using specific hypotheses developed from previous research evidence.

#### **Feasibility of a choice experiment**

The feasibility of the choice experiment approach was assessed by establishing how many patients who were able to take part were willing to do so, whether they said they understood the instructions given and made a choice between scenarios presented to them, and whether they completed the task. A further measure of feasibility was to establish whether respondents traded between attributes and whether the model produced coefficients for the attributes that were significant at the 10% level, and to estimate marginal rates of substitution. It was also important to establish whether the study design with less than 100

participants per sub-group would produce results that would be interpretable and useful for decision-makers, since the previous PDC study had shown that it was hard to recruit more respondents in a reasonable length of time.

#### Hypotheses to be tested in choice experiment

The hypotheses were that specific attributes would be more important to respondents than others and that particular pre-defined groups of patients would have stronger preferences for particular attributes of the service. The justification for these putative associations comes from previous descriptive research in PDC (e.g. Faulkner et al. 1993, Douglas et al. 2000).

Specifically the hypotheses tested in the data were the following:

1. Open access (unstructured day, no appointments), staying all day and therapeutic interventions is positively and significantly related to choice of scenario.
2. Access to a doctor will not be dominant over all other attributes of PDC.

In the sub-group analysis:

3. Patients living alone will have stronger preference for staying all day as they are more likely to suffer from acute social isolation
4. Elderly patients (over 75 years) will have stronger preferences than others do for 10am to 3pm and for staying all day rather than having an appointment.
5. Younger patients (under 65 years) will have stronger preferences for therapeutic interventions (a more active form of intervention).
6. Preferences for different attributes of PDC will depend on the type of centre the respondent attends and the services that are available. (Respondents will value attributes they have not experienced lower than those they have experienced).

#### **Design of the palliative day care choice experiment**

A comprehensive checklist for designing and undertaking choice experiments in health economic evaluation had not been published at the time of designing the PDC experiment. This section presents the methods adopted in the PDC choice experiment and discusses the reasons behind the decisions to follow the example of other studies, or to take a different approach. Since the methods have evolved rapidly over the last 10 years, there is some heterogeneity of study design in the literature and judgement needs to be taken on the most appropriate approach within the context of each health care setting. Where an approach has



been used that is different from the usual (as set out in Ryan 2001b), justification of this decision is presented, along with discussion of the possible consequences of this approach for interpreting the findings of the study.

### **Identifying the attributes and number of attribute levels**

Attributes for the PDC study were selected on the evidence of in-depth observations and interviews undertaken by HRD of palliative day care as described in chapter 3 (Douglas et al 2000). The preparation for the North Thames study had involved HRD visiting all five of the PDC centres who took part in the study for a period of one week consecutively. Following on from this experience, further interviews were held with key members of staff involved in delivering and managing the service, as well as informal discussions with patients while they attended the day centre (HRD worked as a volunteer in each centre and talked to patients during the course of this work). This observational and interview data provided the first idea that a choice experiment might be an innovative approach to exploring the value of specific aspects of the service.

The starting point for defining the attributes for the study was the differences in how palliative day care centres were organised and the emphasis on contrasting aspects of care and what was believed to be the important aspects of palliative day care. Attributes representing more social, open access configuration of services were contrasted with attributes that defined themselves as therapy-based services structured around appointments. Also included were characteristics representing “personal care,” such as hairdressing and bathing. This is because there was no common agreement about whether these should be included in the “package” of services and there were strong views amongst providers about whether these forms of care were appropriate in specialist palliative day care settings. In the North Thames study, some centres offered these services to all patients, while others did not offer them at all. Table 5.1 shows the range of different services offered in the four centres that participated in the choice experiment study.

None of the centres were exactly the same and all provided different activities, depending on the philosophy of care that determined which aspects of care should be emphasised (table 5.1). Furthermore, the comments by the patients who gave their views indicated that patients might have different views from the health care professionals as to what was important about PDC. It was decided that, from a policy-makers’ point of view, the focus should be on the value of those services that were different between centres, such as the timing of and mode of access to care, and particular activities emphasised in some centres and not others (such as access to medical support).

Some of the attributes defined in this early phase of the study were the opportunity to meet other patients and socialise with others facing similar problems, to have a safe place to talk about living with illness, to learn a new skill and try out creative activities, to go on trips outside the centre and be active, and to provide a break for home carers. However these particular attributes were seen as either too ill-defined, or meaning different things to different people. This method of defining the attributes in a way that is relevant to both users of the service and to the decision-makers is common in designing health economic discrete choices (Ryan, Bate et al 2001b).

**Table 5.1 Breakdown of activities and structure of the day, by centre**

	Centre 1	Centre 2	Centre 3	Centre 4
Social activities	•	•	•	•
Doctor appointments	•			•
Nurse-led clinic	•			
Arts activities	•	•	•	•
Music Therapy				
Physiotherapy	•	•	•	•
Hairdressing			•	
Counselling/ Psychotherapy	•	•	•	•
Massage	•	•	•	•
Bath	•			
Hydrotherapy				•
Reflexology	•	•	•	•
% patients who stay in PDC from the beginning to the end of the day.	100%	93%	93%	33%

It was decided to concentrate on service attributes rather than these less well-defined characteristics of palliative day care. This has meant that the study has focussed on service provision rather than the utility that patients may derive from aspects of the service. This means that the study may not have captured important attributes in individuals' utility function. This may be a limitation of the study and is discussed further in the final chapter.

Furthermore, the choice experiment reduced the attributes down to six specific characteristics. This was a limited range of attributes that could describe PDC (which is usually described as more than the sum of its discrete parts) and this is also a limitation of the study. It may not reflect the aspects of the service that patients themselves have described as being the most important aspect of day care, for example getting out and meeting others in similar circumstances (Goodwin et al 2002b). Even if the most important aspect of PDC is that it provides a safe environment to meet others, this does not provide insights into the value of specific services. Also, all centres aim to provide care in an

environment that promotes sociability, and is comfortable and relaxed. If this attribute were included and a scenario was presented that deliberately left out “friendly social atmosphere,” this might reduce the believability, and hence the validity of the experiment. Further, there would be a strong chance that respondents would express dominant preferences for this attribute and no useful data would be collected on the value of other aspects of care.

The attributes are characteristics of the different approaches to palliative day care rather than representing the complete picture of the service. Aspects of care such as “existential health” or spiritual well being have not been addressed in this study. It was felt that the complexity of trying to describe in simple terms a highly complex domain of quality of life was beyond the limitations of this study. Efforts have been underway to explore aspects of existential health in the quality of life literature but this is still embryonic and has been developed for specific contexts that may not be appropriate for palliative day care settings. This issue is explored further in chapter 7.

Social care is represented by hairdressing, and personal care is represented by bathing to explore whether PDC attenders value social and personal care as much as the specialist services that are offered. Centres with a more medical philosophy are represented by routine access to a doctor and centres that only offer specific appointments.

Although the PDC experiment did not include a price or charge to patients as an attribute, other variables that could act as proxies for a monetary valuation were considered. Travel time was not an important consideration for PDC attenders since volunteers (or taxi services) are employed to bring them to and from the centre. Furthermore, it was assumed that for individuals who had almost no other health and social care input and do not participate in other outside activities (as the findings of the North Thames study showed), the opportunity cost of time could be relatively low. This is paradoxical since this group of patients may have only months or weeks to live. But on a day-to-day basis, time is in relative abundance. Therefore a longer journey to a PDC centre could not be assumed to be a proxy for monetary sacrifice and it might even be something that people might look forward to.

Waiting time for a place at a PDC centre was not a useful proxy for monetary sacrifice as none of the centres had a waiting list. Quality of life variables were too complex to consider in this study design. Descriptions would be open to difference of interpretation by respondents and potentially represent different qualitative characteristics of care to individuals. Also, some of the descriptive variables were thought to be too specific. For

example, the utility (or disutility) of a PDC centre that opened for an hour longer or shorter might not be very relevant to their decision to attend.

Therefore attributes that were chosen were fairly simple and clearly demarcated. The identification of more complex attributes that reflect patients' preferences and mirror their decision-making presented in ways that make sense to them is an area of research that needs be explored further, and this is considered in the discussion.

### **Number of attributes**

Six attributes were finally selected in the PDC study (table 5.2). This was a balance between the range of attributes that could have been used in the study and the need to avoid information overload. Each attribute represents a domain or dimension of palliative day care that is 'discrete' (is not related to the presence of any other attribute), is recognisable to most patients who attend day care, and is a variation of day care. This means there are also attributes where there is no agreement (or evidence) on effectiveness or strength of patients' preferences.

To continue to keep the number of scenarios in the choice experiment manageable, it was decided to define the attributes in only two categories rather than a higher number of variations (levels). For the activities (bathing, hairdressing, specialist therapies, and medical support) the attribute was either 'present' or 'absent'. For the attributes that defined the way that the centre worked (opening hours, and structure of the day) these were harder to determine as attribute levels. In-depth discussions with providers of PDC revealed that the usual 10am to 3pm opening hours (for all centres) had been identified as the most convenient for staff (allowing for ward rounds before and after opening hours) but that this might not be the best for the patients or their carers. Originally, three opening times were specified: 10am to 3pm, 1pm-6pm and 3pm to 8pm. Pilot questionnaires tried out on PDC patients showed that the 3pm to 8pm timing was not favoured by any patients and was consequently dropped from the final design.

For six attributes, each with two categories, a full factorial approach would have given  $2^6$  combinations, a possible 62 scenarios to put into choice sets. This is still too many to present to respondents in a questionnaire, despite the simplification of the levels. Three studies have been published in the health economics literatures that have undertaken a full factorial design (Ryan and Farrar 1994, van der Pol and Cairns 1998). Many studies adopted the fractional factorial design. A fractional factorial design was therefore adopted. No interactions between attributes were assumed in the first instance. It would be preferable to

include interactive effects although the sample size would increase substantially to do so. The assumption of no interaction effects is a strong assumption that was explored by adding interaction terms (discussed further on).

### **Number of choice sets and the balance of the study**

Eight pairwise choice sets were specified for this study design. Since six attributes were used to estimate the model, this allowed for sufficient error degrees of freedom (as specified in chapter 4). The minimum number of choice sets was determined using SPSS software version 10, using the program Orthoplan.

One of the determinants of the number of pairwise choices is the decision of how to pair up the scenarios. This study lies at the low end of the range of 8-18 choice sets that has been adopted by other choice modelling studies in the review (Ryan and Gerard 2001). A non-choice ('choose neither') option was not included in this study as the purpose of the experiment was to determine the relative importance of each attribute, and not primarily to consider whether the overall package of care offered in each scenario would be chosen in a real life context. By not including a non-choice option, respondents were 'forced' to make difficult choices between bundles of attributes that they may not have chosen in real life. Since the number of pairwise choices was restricted, it was felt that to offer a non-choice option would further dilute the amount of data collected in the experiment. The result of this is a regression model that has reduced external validity (to reflect actual choices of PDC centre in real life), but increased in internal validity (measuring what it sets out to measure in terms of the relative value of the individual attributes).

The issue of design balance and orthogonality was addressed by adhering to recommendations that choice experiment designs use a common ratio for the number of levels of attributes (Ryan, Bate et al 2001). As there is no status quo in palliative day care in the real world it was felt a design with a status quo scenario to compare all others was conceptually flawed. When all scenarios are compared to one constant only, no empirical data can be collected on the comparison of relative preferences between all other scenarios that are not the constant. The results are dependent on and limited by the choice of constant. This could be a limiting aspect of this type of study design.

What happens to the statistical properties and balance of the design when the scenarios are put into pairs either randomly or against one status quo scenario has not been explicitly described in previous studies. Without a recipe to follow (and not wanting to follow one blindly), this study took a slightly different approach. Assessing the problem from first

principles, it has been established that choice experiments model the difference between the attributes in each pair, that is, the change in welfare or utility of moving from a scenario where an attribute changes by one level. Consequently, it might make sense to consider the orthogonality of the design related to the difference in levels of attributes between pairs. In other words, the difference between the levels of two attributes should be computed in the orthogonal design, not the levels themselves. This design principle of modelling the difference between attributes was also adopted by Ryan and colleagues in a recent study (Ryan et al 2000). In this study, it was decided that the pairs of choices should be orthogonal rather than the scenarios themselves since creating an orthogonal design and then dividing these scenarios into choice sets would potentially compromise their orthogonal property. Without any clear reasoning from first principles as to why we should not take this approach we set about designing the orthogonal array of pairs of scenarios as follows.

In a design where only two levels of an attribute exist, four possible combinations of pairs of attributes are possible. For example, the combinations for the attribute “specialist therapies” could be:

- w) Specialist therapies are available in centre A but not in centre B
- x) Specialist therapies are available in centre B but not in centre A
- y) Specialist therapies are available in both centres
- z) Specialist therapies are available in neither centre

Each of these combinations were coded W, X, Y, Z. It was necessary to trade-off the maximum number of questions that could be presented to each individual while still gathering enough data to run the analysis. Therefore it was proposed that only pairs W and X, which presented differences in attributes between scenarios, should be included in the final experiment. The reason for this was that including pairs Y and Z in the orthogonal array would substantially increase the numbers of possible combinations that could be generated. It is argued that if the level of an attribute is the same in both scenarios in a pairwise choice (both offer specialist therapies), then this attribute would not contribute to the choice between them.

Six attributes with two combinations (present in centre A or present in B) would generate 64 possible scenarios. Adding two more combinations to the choice (an attribute level present in both, or present in neither) would generate 256 scenarios. For a fractional factorial design with only eight pairwise choices per respondent, the restricted design without any attributes the same in scenario A and B covers a higher proportion of all the possible combinations of

pairs of scenarios. Therefore only the codes W and X (where the level of attribute was different between scenarios) were entered into the orthogonal design. The consequence of this is reviewed in the discussion of the findings.

A discrete choice model was generated using codes W, X with eight pairs of scenarios (SPSS 10.0, Orthoplan) providing information on which attribute should appear in which scenario. These pairs of scenarios were balanced and orthogonal, that is, each pair of scenarios was not correlated with any other pair.

**Table 5.2 Final attributes of PDC included in the discrete choice model**

PDC Attribute	Description of the attribute and the model measures	Choice set	
Access	Centres that encourage people to stay all day versus centres where patients attend for an appointment (1-2 hours) and then go home.	Appointment-based	Stay all day
Timing	Centres open at different times of the day	10am-3pm (current)	1pm-6pm
Massage, reflexology, aromatherapy	Physical/caring/therapeutic activities designed to improve patients' physical symptoms, body image, self-confidence etc	Specialist therapies available	Not available
Doctor	Routine medical consultation as part of the basic PDC package	Available by appointment every visit	Only seen in emergencies
Bathing/ Hydrotherapy	Access to baths at the PDC centre	Available	Not available
Hairdressing	Hairdressing as a form of personal and social care	Available	Not available

### The pilot study

Once the number of choices and range of attributes had been agreed, a pilot phase of the project was undertaken. This took place in two stages. The first stage was to administer the questionnaire to a convenience sample of LSHTM public health students to assess whether the instructions were understandable, whether the task could be completed (level of difficulty) and to find out any comments they had on the process. The second was to test the final questionnaire on PDC patients.

In the first phase, students who were willing to take part stayed behind at the end of a lecture and were asked to fill in a self-completed questionnaire with eight choice sets. They also filled out two extra questions about the task at the end. This included a visual analogue scale from 1 (extremely easy) to 10 (impossibly difficult) and an open ended section for comments.

Twenty-three students began the task and 19 returned the form with all eight choices completed. The average score on the visual analogue scale of difficulty in completing the task was 7.4, which was higher than had been expected. Eleven students provided reasons for their answers. The comments were mostly negative. Of the students who scored it over 8 (six students), comments were: "I do not think palliative care patients will understand it, especially if they're really ill"; and "There are too many things going on at once - too hard to understand each of the variables at the same time. Exhausting!"; and "Really hard to judge what patients want. Do they really want hairdressing? Cheap I suppose!"

In light of these results, the idea that the questionnaire should be administered as a face-to-face interview was reinforced. Further meetings were held with PDC managers who also expressed serious doubts as to whether patients would be able to understand and complete the task. Changes were made to the design of the study. First, it was decided that only patients with some experience of the range of activities offered in palliative day care services should be interviewed since those who did not know the service found it hard to value its attributes. Only patients who had attended for at least one month would be eligible for interview in the main study.

A second pilot study was undertaken in the Chichester hospice while the EQ-5D study was under way and before the main choice experiment study was undertaken to assess the feasibility of the questionnaire and interview process with PDC patients. The choice experiment questionnaire was administered to four patients (two women, two men). An informal interview was undertaken after each questionnaire to find out whether they had understood the task, how hard they thought it was and whether it could be improved.

None of the respondents said they had any problems either understanding or completing the task. While they said some of the choices presented were more difficult than others, they were able to make a choice and complete the task. The concerns they expressed were whether they were being consistent in answering the questions in each paired choice. One of the patients could not see well and so the questionnaire had to be read out to him. He did not say this had been a problem and reported no other difficulties.

All the respondents said they strongly disfavoured the category '3pm - 8pm' a category of the attribute 'Time' and said they would not want to attend at all if this is when PDC was open. One patient wanted to know at the end if the centres were the same in other respects other than these particular attributes. The completion time for the questionnaire was in the range of 10 - 20 minutes.



On the basis of this pilot, it was assumed that people who knew the service would be more enthusiastic about taking part than people who did not (that is, asking members of the public to value PDC). Second, the 3pm-8pm category was dropped from the experimental design. Third, the text explaining the task emphasised more strongly that the centres were to be assumed to be exactly the same in all other respects other than the difference in attributes.

### **Sample size**

Data were collected using face-to-face interviews. It was not thought that postal questionnaires would yield complete responses, and there was a danger that patients would not understand the task or would feel obliged to answer the questionnaire. Also, the small number of patients attending PDC in any week was one of the reasons why a face-to-face interview was considered to be a better format for data collection. Also, given the level of ill-health and age of the patients, it would seem appropriate to take more time to make sure that respondents understood the task and to ensure that completion of the task was as high as possible.

To assess how many respondents needed to be recruited, the analysis was undertaken in a number of stages. When 25 respondents had been interviewed the first model was estimated, then for 50 respondents, then 75. The purpose of this was to assess whether the coefficients on any of the parameters in the model (the attributes of the service) were significant at each of these stages, and whether this changed or remained stable as more data were added. If it appeared that the addition of data from more respondents was not improving the model, and, as importantly, would not change the conclusions of the study, then recruitment to the study would stop. This is similar to qualitative approaches to sampling where interviewing continues until a point where no additional information is gathered from additional data collection (Kirk and Miller 1985). It was considered be unethical to continue to recruit patients to a study when their responses were unlikely to change the findings and consequent decisions about how services ought to be organised.

It was foreseen that it would be necessary to ensure a high completion rate since the total number of people attending PDC (the pool from which respondents could be recruited) across all four centres was not large. The number of patients registered to attend PDC at the time of data collection was as follows: Centre A 80 patients; Centre B 35; Centre C 40; Centre D 30. Centre A reduced its opening days from three to two days per week due to lack of referrals while data collection was on-going.

This number is misleading as a proportion of patients will not attend on any given week. The reasons given for why people do not attend were that they are away on holiday, attending a hospital appointment, or too unwell to attend. The first multi-centre study found that patients attended, on average between two and four times per month. In the centre where patients were encouraged not to stay for the full day, it was also likely that some patients would turn up for one appointment and leave immediately. There was some concern that a proportion of this group of patients would be missed by the interviewer. However, the majority of patients attended for more than one appointment on any given day, so stayed for half a day, morning or afternoon. They were interviewed between appointments.

### **Eligibility criteria**

Patients were only approached after the PDC leader had decided that patients were well enough to be interviewed. Patients had to understand English sufficiently to be able to read (or have the questions read to them) and understand the descriptions of the attributes. Physical disability was not a barrier to taking part: patients who could not sign their consent form were included and their consent form signed and witnessed by two volunteers in their presence. There were no other exclusion criteria.

### **Scenario presentation and data collection**

Scenarios were presented to patients as descriptions of two PDC centres that they could attend (see an example of the choice scenario in Appendix C). Data were also collected on respondents' age, whether they lived alone, how long they had attended PDC, and what services and activities they usually accessed in PDC. It also asked them to say what aspect of PDC was most important to them. Notes were taken contemporaneously on whether patients said they had dominant preferences for a particular attribute or combination of services.

### **Analysis of the data**

A computer package specifically intended to design and analyse the results of choice experiments was not used in this study. Instead the approach taken was to use a generic statistical package (SPSS 10.0 to generate the orthogonal array and STATA 7.0 to run the parameter estimation model). The reason for this was that the approach taken in designing the experiment was different to that set out in the published literature as no pair of scenarios contained attributes that were the same levels. Since probit and logit models can be generated in generic packages and the steps can be clearly delineated, this was the approach adopted. It enhanced the transparency of the methods used.

An ordinary probit, fixed effects and a random effects probit model were estimated using STATA, assuming in the first instance independence between attributes as well as absence of correlation between covariates and individual effects. As each individual was asked to make a series of pairwise choices, the error terms could not be assumed to be independent since two observations made by the same person are more likely to be correlated than two observations from different people. Therefore a panel data technique was adopted. As the option to “choose neither” of the two scenarios was not offered, an unconditional demand curve was estimated. This approach is consistent with the most recent guidance published in reviews of methods for health economic evaluation using choice experiments (Ryan and Gerard 2001, Ryan, Bate et al 2001).

As all attributes were binary, the probit model was structured so that if a specified level of an attribute were present in scenario A (absent from scenario B), this would be represented by 1. If it were absent from scenario A (present in scenario B), this would be represented by 0. The attribute that was present or absent, or the level, is listed in table 5.3.

**Table 5.3. Specification of attributes and levels for the probit / logit model**

Attribute	Level =1	Level =0
Access	Stay for the full session	Appointment only
Time	1pm – 6pm	10am – 3pm
Bath	Available	Not available
Specialist therapies	Available	Not available
Hairdressing	Available	Not available
Doctor	Available every visit	Emergencies only

The difference between scenarios was coded as scenario A minus scenario B. If a pairwise choice had “Stay for a full session in A” and “Appointment only” in B, then this was coded 1 (=1-0). The converse was coded -1 (=0-1). For this study, the following equation was estimated:

$$Y = \text{constant} + \alpha_1 \text{ access} + \alpha_2 \text{ time} + \alpha_3 \text{ bath} + \alpha_4 \text{ specialist} + \alpha_5 \text{ hair} + \alpha_6 \text{ doctor}$$

with Y representing the change in benefit of moving from scenario B to scenario A, and  $\alpha_{1-6}$  representing the difference in utility represented by the presence/absence of each attribute.

A significant constant term indicates that respondents have a predisposition (or bias) towards choosing scenario A or B, all attributes being equally present or absent. The constant term might be significant if, for example, one scenario represented the status quo,

or if respondents always chose the scenario on the left hand side of the page. An insignificant constant term indicates no such bias.

### **Attribute/covariate interactions**

The interaction between attributes in the choice experiment and specific covariates (characteristics of the respondents and how they accessed palliative day care) were examined. It was hypothesised that these covariates might have an impact on respondents' value of different kinds of services. Age was considered to be an important factor in people's preferences for different models of palliative day care. Other covariates relating to the circumstances of the respondents, such as whether they lived alone, were considered. There is also an argument that the length of time since first attending PDC and the intensity of care (measured by days of attendance per week, or by the number of activities per visit) could also be important. One limitation of the research design was that activity data was only collected for patients attending the Chichester palliative day care centre, so that information on individuals' use of resources was not collected for the majority of respondents in the other three centres participating in the choice experiment. Demographic data and other information on whether the patient lived alone were collected from all respondents.

The following dummy variables were created. They were OLD (if the respondent was 75 years or older at the time of interview); YOUNG (under 65 years); LIVALONE (if the patient lived alone); MEDCENTRE (if the centre the respondent attended had access to routine medical appointments); ALLDAY (if the centre encouraged patients to stay from opening till closing time). Interaction terms were created with these dummy variables. They were: OLDACCESS and YOUNGACCESS to assess whether the age of the respondent had an impact on whether they wanted to stay for the full session; ALONEACCESS to assess whether patients who lived alone had a greater preference for staying for the full session; ALLDAYACCESS to assess whether respondents who attended centres that encourage people to stay all day had a stronger preference for this attribute than patients who attended appointment-based centres; MEDCENTREDR to assess whether patients who attended centres with routine medical appointments preferred this more than those who did not; and YOUNGOTHERAPY to estimate whether younger patients had a greater preference for more active, therapeutic interventions. The probit model was estimated for each of these interaction terms consecutively.

## **Assessing the goodness of fit of the model**

### **Hold-out scenarios**

In addition to the eight pairs of scenarios in the experiment, the orthogonal design programme also produced two hold-out pairwise choices, (choice nine and ten) which the respondents also answered in the same way as choices one to eight). A McFadden pseudo-R<sup>2</sup> statistic was also predicted for the model.

### **Tests of rationality**

One of the important considerations for this study was the trade-off between the number of choice sets that respondents could be reasonably expected to undertake, and the range and amount of information that could be gathered.

In this study, the only test of rationality that was adopted was a simple dominance test. No multiple tests were adopted. The PDC study identified all respondents who expressed dominant preferences. These were respondents who consistently chose the scenario with one particular attribute. Clear agreement on whether these respondents should be removed from the analysis could not be found in the recent literature, with some studies including dominant respondents in the analysis (Shackley et al 2001), others explicitly dropping them from analysis (Ryan et al 1999, Ryan et al 2000), or not specifying which approach was taken (Morgan et al 2000). In the PDC study, the analysis was undertaken with and without these respondents to assess whether this had an effect on the parameters of the model.

One of the possible limitations of this study design in assigning only two levels of each attribute was that a high number of false dominant preferences might be detected. This is because the range of choices open to respondents is more limited than for experiments that use three or more levels. Less choice will mean potentially more respondents making decisions “as if” they had dominant choices when in fact more choice would have indicated the point at which they would be willing to trade between attributes.

### **Estimating the importance of attributes**

Since no cost attribute was included in the study, it was not possible to calculate a marginal WTP for any of the attributes. This way of reporting the results still remains controversial. What is more important is that the six attributes that are included in the experiment are all qualitative categorical variables. There is no continuous variable against which to compare all other attributes. Such an attribute could be travel time or waiting time. Since no such proxy for cost (or rather sacrifice) was used, there was no method of comparing all attributes in a way that would demonstrate the sacrifice respondents were prepared to make in order

to have a specific attribute in a PDC centre. This is another important limitation of the study. The limitations of the study design will be addressed after the results are reported.

## **SUMMARY**

The study was designed so that the choice experiment followed on from the EQ-5D data collection period and did not overlap with it. The EQ-5D was fairly straightforward. Three groups of patients were identified, a before group, a contemporary group not receiving PDC and a group attending PDC. The study followed up patients attending or not attending PDC for a three month period. This provided the opportunity to measure changes in quality of life using a more robust study design than had been adopted in the North Thames palliative day care study. However, since no difference in length of life was expected, the null hypothesis was that EQ-5D would not detect significant differences in quality of life.

The choice experiment identified six attributes of PDC that were different across centres and where the relationship between the specific component of the service and quality of life was not well understood. Methods of undertaking choice experiments had not been uniformly reported and this study was an attempt to design an experiment from first principles. Since the study was first planned, other health economic studies have reported their methods of analysis more widely, and the methods used here have been compared with these. Where the methods have varied, a detailed discussion of the study design used here has been shown.

The next chapter reports the findings of the EQ-5D study and the choice experiment.

## **Results of the empirical research into palliative day care**

### **Introduction**

The results of the Chichester Palliative Day Care study ('the Chichester study') are presented here, although not all the data that were analysed for this study are reported in this thesis. The before-and-after study gathered data on the cost of care and used the same quality of life instruments used in the North Thames study. The data on the disease specific outcomes, The Palliative Outcome Scale, and the Herth Hope Index were analysed by other researchers. They reported that no differences in quality of life could be determined between patients who attended PDC and those who did not, either in the historical or contemporary comparison groups (Higginson, Goodwin et al 2002) These results are pertinent to the interpretation of the findings of the EQ-5D analysis in this chapter. The choice experiment analysis was undertaken after the analysis of the quality of life outcomes was complete. Results are reported in a similar way to the reporting of choice experiment results published recently in the health economics literature.

### **SECTION 1. DATA COLLECTION AND FOLLOW-UP**

The 'before' group of patients were recruited over two months from all patients within the palliative care service at a hospice in Chichester (inpatient and home care nursing). As described in the previous chapter, these were patients whom the clinical team believed would benefit from attending palliative day care. Once the centre had been built, the day care group and a contemporary comparison group were recruited over fifteen months. During this time, there were 63 new referrals to PDC, of which 40 consented to participate. The reasons for those who did not participate were: refusals (n=7), too unwell (n=10), brief attendance for out-patient consultations only (n=6).

In all, 140 people were recruited to the study. Table 6.1 shows the recruitment to each patient group and completed interviews (percentages are given as a proportion of total recruited). There was more attrition in the PDC group before first interview compared with the before and comparison groups. The main reasons for attrition in all the patient groups

were death and feeling too unwell. Only one patient refused to be interviewed in the PDC group.

**Table 6.1 Completed patient interviews in the EQ-5D study**

	Before group (n=82)	PDC group (n=40)	Comparison group (n=49)
Not interviewed	16 (20%)	12 (30%)	3 (6%)
<b>Completion rates:</b>			
1st interview	66 (80%)	28 (70%)	46 (94%)
2nd interview	36 (44%)	16 (40%)	30 (61%)
3rd interview	26 (32%)	11 (28%)	20 (41%)
Reasons for no interview	Death (n=28) Too unwell (n= 9) Refused (n= 7)	Death (n= 7) Too unwell (n=15) Refused (n= 1)	Death (n=10) Too unwell (n=15) Refused (n= 4)

The patient groups were analysed in sub-groups defined by how many times they were interviewed. Table 6.2 below shows the number of patients in each group.

The palliative day care group had the greatest drop-out of patients over three interviews (from 28 at the start to only 11 patients by third interview). Therefore the cost data and EQ-5D analysis must be interpreted with some caution since the group is so small by third interview.

**Table 6.2 Numbers and proportions of respondents with completed interviews for each data collection stage (grouped by intention to treat)**

Interviews completed		Number of respondents
Before group	Only 1	30 (45%)
	1&2 at least	36 (54%)
	1,2&3	26 (39%)
<b>Total</b>		<b>66</b>
PDC group	Only 1	12 (42%)
	1&2 at least	16 (57%)
	1,2&3	11 (39%)
<b>Total</b>		<b>28</b>
Comparison Group	Only 1	16 (35%)
	1&2 at least	30 (65%)
	1,2&3	20 (43%)
<b>Total</b>		<b>46</b>



The socio-demographics in the patient groups were similar (chi-square  $\chi^2$ ,  $p \leq 0.05$ ) suggesting that it was reasonable to make comparisons between groups. Table 6.3 shows the patient socio-demographics for those that completed the baseline interview. The age of the groups is reflected in the employment status, as the majority of patients were retired (over 70% in each group). There were no differences between the groups in gender, ethnicity or type of carer contact.

Table 6.3 Patient socio-demographics (completed 1st interview)

Demographics	Before group (n=66)	PDC group (n=28)	Comparison group (n=46)	Significance
<b>Age in years</b>				
Mean (SD)	69.2 (12.4)	74.0 (10.1)	70.8 (11.9)	p=0.197
Median/ range	71.0/ 34-94	77.0/ 50-94	72.0/ 39-90	
<b>Gender</b>				
Female	40 (61%)	12 (43%)	23 (50%)	p=0.24
Male	26 (39%)	16 (57%)	23 (50%)	
<b>Ethnicity</b>				
White UK	66 (100%)	28 (100%)	46 (100%)	
<b>Employment status</b>				
Working (F/T or P/T)	6 (9%)	2 (7%)	2 (4.5%)	p=0.45
Not working (unable)	14 (22%)	2 (7%)	9 (20.5%)	
Retired	45 (69%)	23 (85%)	33 (75%)	
<b>Carer</b>				
Spouse	43 (65%)	20 (71%)	30 (70%)	p=.067
Other carer	12 (18%)	5 (18%)	9 (21%)	
No carer	11 (17%)	3 (11%)	4 (9%)	
<b>Carer contact</b>				
Lives with spouse	43 (65%)	20 (71%)	30 (70%)	p=0.58
Lives with family	2 (3%)	0	2 (5%)	
Lives alone	21 (32%)	8 (29%)	11(25%)	
<b>Carer employment</b>				
Working (F/T or P/T)	18 (29%)	6 (21%)	8 (19%)	p=0.67
Not working (unable)	5 (8%)	2 (7%)	4 (9%)	
Retired	29 (46%)	17 (61%)	27 (63%)	
No carer	11 (17%)	3 (11%)	4 (9%)	
<b>Housing</b>				
Private ownership	28 (42%)	11 (39%)	17 (37%)	p=0.73
Council housing	7 (11%)	7 (25%)	7 (15%)	
Rented housing	28 (42%)	9 (32%)	20 (44%)	
Other	3 (5%)	1 (4%)	2 (4%)	
(e.g. nursing home)				
<b>Primary diagnosis</b>				
Cancer site:				p=0.58
Lung	11 (17%)	4 (14%)	11 (26%)	
Gastrointestinal	11 (17%)	8 (29%)	9 (21%)	
Breast	9 (14%)	4 (14%)	4 (10%)	
GU / Prostate	11 (17%)	6 (21%)	8 (19%)	
Gynaecological	7 (11%)	0	3 (7%)	
Other cancer site	10 (15%)	4 (14%)	7 (17%)	
Non-cancer diagnosis	6 (9%)	2 (7%)	0	
<b>Place of death</b>	<b>(n=46)</b>	<b>(n=11)</b>	<b>(n=22)</b>	
Home	8 (17%)	4 (36%)	6 (27%)	p=0.69
Hospital	7 (15%)	1 (9%)	3 (14%)	
Hospice	31 (67%)	6 (55%)	13 (59%)	

## SECTION 2. ANALYSIS OF THE COSTS ASSOCIATED WITH PDC

### Summary of overall costs of care

Tables 6.4, 6.5 and 6.6 show the mean (standard deviation) costs for one month's care between the three groups, over the three time points that data was collected during the study. The standard deviations are very large and the differences in costs between the greatest and the least are also large. This demonstrates that costs of one month's care differ widely between patients within palliative care groups (palliative day care, before group and the contemporaneous comparison group).

**Table 6.4 Cost of four weeks' care: patients with only the first set of data at baseline**

Patient group	Time 1 only
PDC group	£888 SD >1000
Comparison group	£711 s.d.550
Before group	£1570 SD >2000

**Table 6.5: Cost of four weeks' care: patients with data for Interview 1 and 2**

Patient group	Time 1	Time 2
PDC group	£720 SD 960	£869 SD >1000
Comparison	£602 SD 882	£505 SD 844
Before group	£412 SD 561	£609 SD >1000

**Table 6.6: Cost of four weeks' one month's care: patients with data for interview 1, 2 and 3**

Patient group	Time 1	Time 2	Time 3
PDC group	£678 SD >1000	£541 SD 624	£492 SD 606
Comparison group	£642 SD >1000	£633 SD >1000	£721 SD >1000
Before group	£384 SD 602	£381 SD >1000	£506 SD >1000

Analysis of the patterns of costs was undertaken using the analysis of variance (ANOVA). This method tests whether the means from two or more data sets are equal (that is drawn from populations with the same mean). This is a development from the t-test which is a test for two sample means only. The p-values provides evidence of the likelihood that the data from the three groups indicate significantly differences in patterns of costs, rather than this occurring by chance. Table 6.7 shows that at none of the time points do the mean costs of care differ significantly (do not reach 10% level of significance, which is a

low threshold for significance). This is due to the small numbers in each cell, as well as the widely different costs within each group.

**Table 6.7: P-values from Analysis of Variance (ANOVA) tests of difference in mean costs of care at time 1, 2 and 3, by palliative care group.**

Patient group (completed interviews)	Time 1	Time 2	Time 3
All patients	p=0.72		
Patients completing at least 2 interviews	p=0.33	p=0.60	
Patients completing all three interviews	p=0.46	p=0.96	p=0.49

The analysis of variance for changes in costs between palliative care groups did not show any evidence of systematic differences between the groups. Figure 11 and 12 show this graphically.

The histograms below represent the means of costs for patients in each group over the three time periods. They must be interpreted with caution as the previous analysis has shown that none of these observed differences is significant. It shows the patterns of costs over time for the group of patients sampled but the likelihood that this pattern would be observed with a second sample is very small.

**Figure 1: Mean Costs for patients in each group over the three time periods**

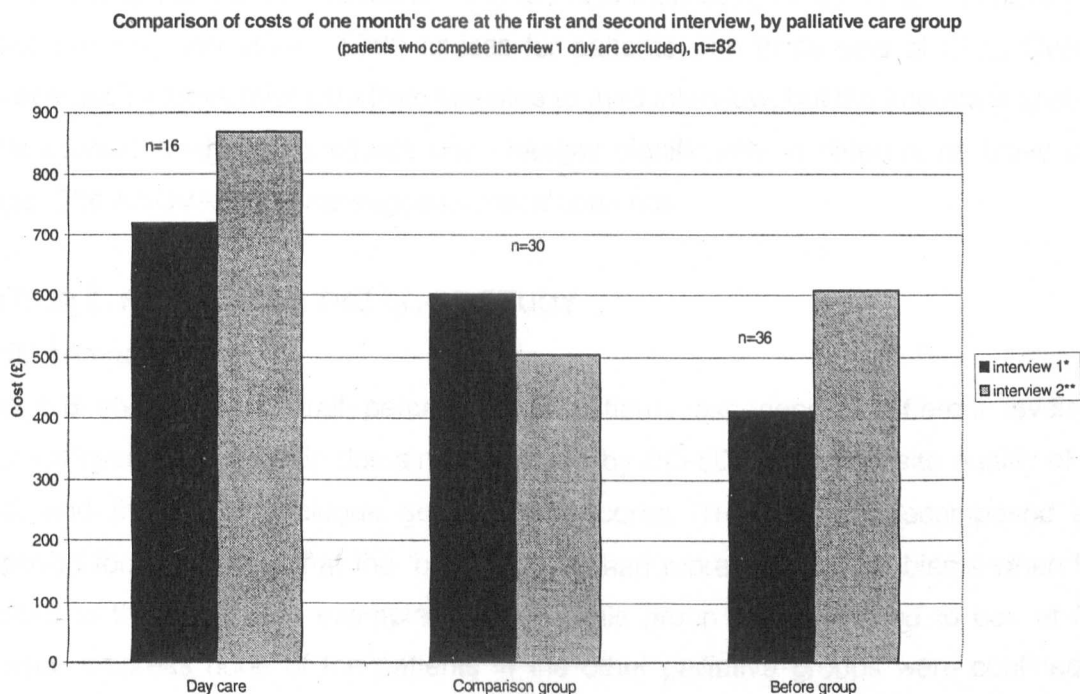
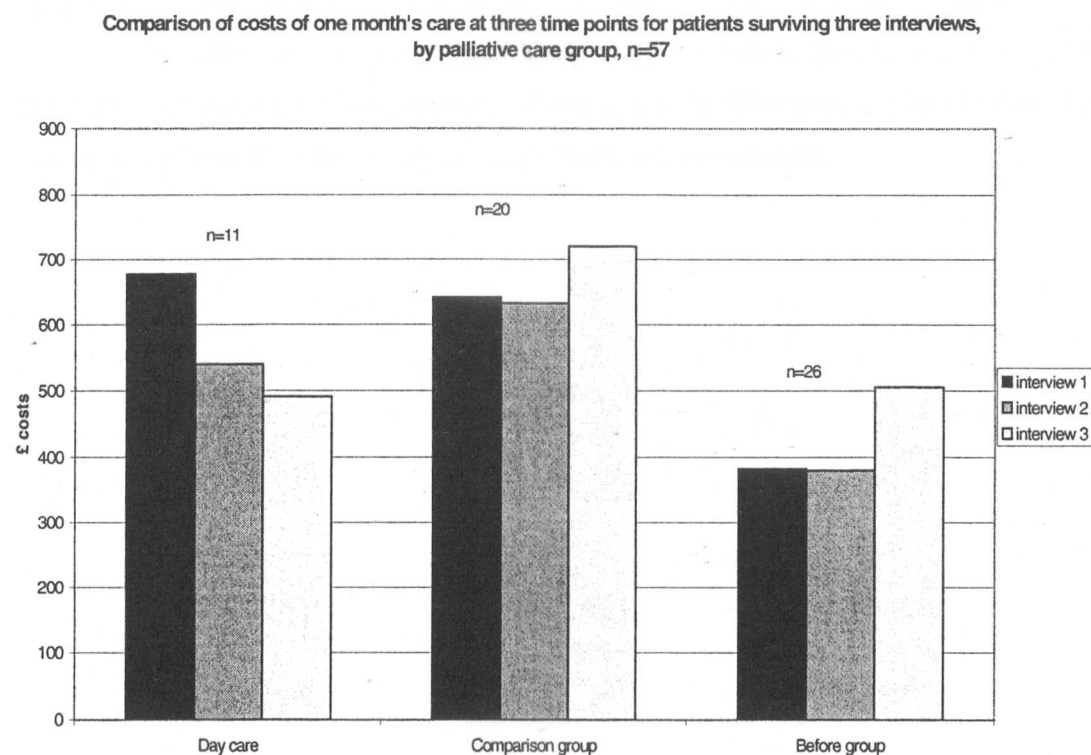


Figure 2: Means Costs for patients in each group over the three time periods



The insights from the cost data are limited since they were collected from one PDC centre only (in Chichester). Overall, it would appear that patients in all three groups are not accessing large amounts of expensive health and social care resource. The total costs of care (including community and acute care) are less than £900 for patients who have data for two periods, and around £700 or less for patients with three sets of data. Overall, resource use reduces over time from baseline to third interview, but the important analysis in this context is whether resource use changes significantly in different palliative care groups. The ANOVA evidence suggests that it does not.

## SECTION 2. RESULTS OF THE EQ-5D STUDY

### EQ-5D domains

Table 6.8 shows the overall percentage of patients experiencing different levels of problems over the five basic domains measured by EQ-5D, the composite quality of life score, and the visual analogue scale (VAS) scores. This data in decomposed and composed form suggests that the 'before' group had more extreme problems when first recruited to the study. For example, 6.1 % of this group were confined to bed at first interview, whereas none of the patients in the other palliative groups were confined to

bed. For all dimensions, the 'before' group reported a higher proportion of 'extreme problems' than the other palliative care groups. Examining the proportions of patients reporting they experienced 'no problems' across palliative day care groups is insightful. By the third interview (the group most likely to be the fittest, that is, the group of patients who have not declined rapidly or died), there is a wide difference in the number of patients who reported that they experienced no problems with self-care.

Table 6.8 Percentage of patients at level of each domain of EQ-5D, by group, and by interview (int.) 1, 2 and 3, compared with the general UK experience.

INTERVIEW 1		General pop.			PDC group			Comparison group Before group					
		Int 1	Int 2	Int 3	Int 1	Int 2	Int 3	Int 1	Int 2	Int 3	Int 1	Int 2	Int 3
Sample		n=28	n=16	n=11	n=46	n=30	n=20	n=66	n=36	n=26			
<b>Mobility</b>													
No problems (%)	95.1	23.1	31.3	18.2	37	30	20	24.2	36.1	30.8			
Some problems (%)	4.7	76.9	68.8	81.8	63	70	80	69.7	58.3	69.2			
Confined to bed (%)	0.2	0	0	0	0	0	0	6.1	5.6	0			
<b>Self care</b>													
No problems (%)	96.7	53.8	56.3	90.9	69.9	80	60	50.8	47.2	50			
Problems (%)	2.9	42.3	43.8	9.1	28.3	20	35	40	44.4	42.3			
Extreme (%)	0.3	3.8	0	0	2.2	0	5	9.2	8.3	7.7			
<b>Usual activities</b>													
No problems (%)	85.8	23.1	31.3	27.3	28.3	36.7	40	16.9	33.3	19.2			
Some problems (%)	11.7	53.8	62.5	63.6	60.9	56.7	45	55.4	41.7	69.2			
Extreme (%)	2.4	23.1	6.3	9.1	10.9	6.7	15	27.7	25	11.5			
<b>Pain discomfort</b>													
No (%)	49.5	34.6	37.5	27.3	32.6	26.7	45	31.8	38.9	30.8			
Moderate (%)	30.5	61.5	62.5	54.5	58.7	73.3	55	59.1	55.6	65.4			
Extreme (%)	1.5	3.8	0	18.2	8.7	0	0	9.1	5.6	3.8			
<b>Anxiety/ depression</b>													
None (%)	73	42.3	50	63.6	56.5	56.7	70	60.6	61.1	53.8			
Moderate (%)	15.7	53.8	50	36.4	43.5	43.3	30	33.3	36.1	42.3			
Extreme (%)	0.7	3.8	0	0	0	0	0	6.1	2.8	3.8			
<b>VAS score (0-100)</b>													
Mean (%)	85.3	60.38	66.56	63.18	64.67	63.28	64	60.59	64.31	62.31			
S.D.	8.3	15.21	14.8	12.5	19.36	17.49	20.37	21.23	19.65	19.65			
Difference (change between interviews)			-6.18	3.38		1.39	-0.72		-3.72	2			
	Mean composite EQ-5D score*	0.555	0.673	0.575	0.627	0.689	0.669	0.496	0.573	0.605			
	S.D.	0.239	0.149	0.292	0.268	0.177	0.247	0.313	0.334	0.251			

\* This is the value used to weight life years gained and produce QALYs.

### Item-response for the EQ-5D data

The item-response rate was very high. All respondents except one who took part in an interview managed to answer all five questions in the EQ-5D schedule. For one respondent there is missing data for item 2 (self care) for the first baseline interview. QALY scores could be derived for all other respondents for all interviews.

### Validation of the EQ-5D against disease-specific health-related quality of life instruments

Correlations were run for each palliative day care group, at each interview (baseline, second and third interview).

**Table 6.9 Pearson correlation coefficient scores and p-values for the 'Pain' domain on POS and EQ-5D, by palliative care group and by interview**

	PDC	Before	Comparison
Interview 1	0.785 p<0.01	0.660 p<0.01	0.501 p<0.01
Interview 2	0.931 p<0.01	0.674 p<0.01	0.878 p<0.01
Interview 3	0.389 p=0.24	0.481 p<0.05	0.678 p<0.01

The data in table 6.9 indicate that there is a positive correlation between the EQ-5D scores and POS scores for the domain 'Pain' at baseline and first interview. By third interview, the numbers in each cell are small (11 respondents in the day care group) which may be the reason why the correlation coefficients are no longer significant at the 5% level. It indicates that the EQ-5D instrument has some internal validity in the 'pain' domain. This is not surprising since the way the question has been framed in both instruments is very similar.

**Table 6.10 Pearson correlation coefficient scores and p-values for the 'anxiety/ depression' domain on POS and EQ-5D, by palliative care group and by interview**

	PDC	Before	Comparison
Interview 1	0.240 p=0.237	0.513 p<0.01	0.625 p<0.01
Interview 2	0.719 p<0.01	0.533 p<0.01	0.575 p<0.01
Interview 3	-0.289 p=0.389	0.406 p=0.76	0.577 p<0.01

The data presented in table 6.10 above is less easy to interpret since the correlation coefficient for interview 1 (which contains the highest number of responses) for the day care group is very low and is not significant at the 5% level. One of the reasons for this may be that, in palliative care research, anxiety and depression are considered to be separate entities (a person can be anxious but not depressed, and vice versa). Another reason for this lack of correlation may be that respondents considered the questions in POS ('have you been anxious or worried?') to be different from the EQ-5D question (I am not/moderately/extremely anxious or depressed). However, a more likely explanation is that this may be a statistical artefact due to the small numbers in the day care group. For the



other (large sample size) groups, the correlations are significant at the 1% level for baseline and first interview.

### **Composite EQ-5D scores.**

Table 6.8 presented earlier also shows the mean quality of life scores, with their standard deviations. The pattern is more easily shown in figures 3 to 5 (at the end of this section) which are the box plots derived from the EQ-5D data. They show the median, quartiles, and extreme values. The box represents the interquartile range that which includes 50% of values. The 'whiskers' are lines that extend from the box to the highest and lowest values, excluding outliers. A line across the box indicates the median. They show visually that the median values across all intervention groups appear to be similar, and that there is no dramatic change in quality of life over time in any groups (that is, between interviews 1 and 2 or between interview 2 and 3). The similarity of results is greatest for patients who have data for three time points (who may have the fewest physical symptoms).

### **Comparison between the 'before PDC and 'after PDC' groups**

Independent sample t-tests were undertaken to compare mean EQ-5D scores for the two groups of patients, before and after PDC (tables 6.11 to 6.13). The t-test procedure was employed to compare means for the 'before' and 'after' groups. Ideally, for this test, the respondents are randomly assigned to two groups, so that any difference in response is due to the intervention and not to other factors. In this study, randomisation did not take place, but checks were undertaken that established that the groups had similar characteristics. Table 6.3 showed that patient characteristics were not significantly different between groups so that it is reasonable to assume that any differences detected could be due to the effect of attending PDC.

Comparisons of composite scores were made between groups for patients undertaking one interview only, at least two interviews, and for patients completing all three interviews.

**Table 6.11 Mean quality of life score at interview 1, and results of independent sample t-test, for patients completing only one interview (p values reported)**

Interview	Intervention group	N=	Mean	Std Dev	P value *
Quality of life score at interview 1	PDC group	12	0.502	0.211	0.689
	Before group	29	0.467	0.327	

\* equal variances not assumed

**Table 6.12 Mean quality of life score at interview 1 and 2 and results of the independent sample t-test, for patients completing 2 interviews or more**

Interview	Intervention group	N=	Mean	Std Dev	P value *
Quality of life score at interview 1	PDC group	14	.600	.260	.365
	Before group	35	.521	.305	
Quality of life score at interview 2	PDC group		.674	.1490	.141
	Before group		.573	.335	

\* equal variances not assumed

**Table 6.13 Mean quality of life scores for patients who completed all three interviews**

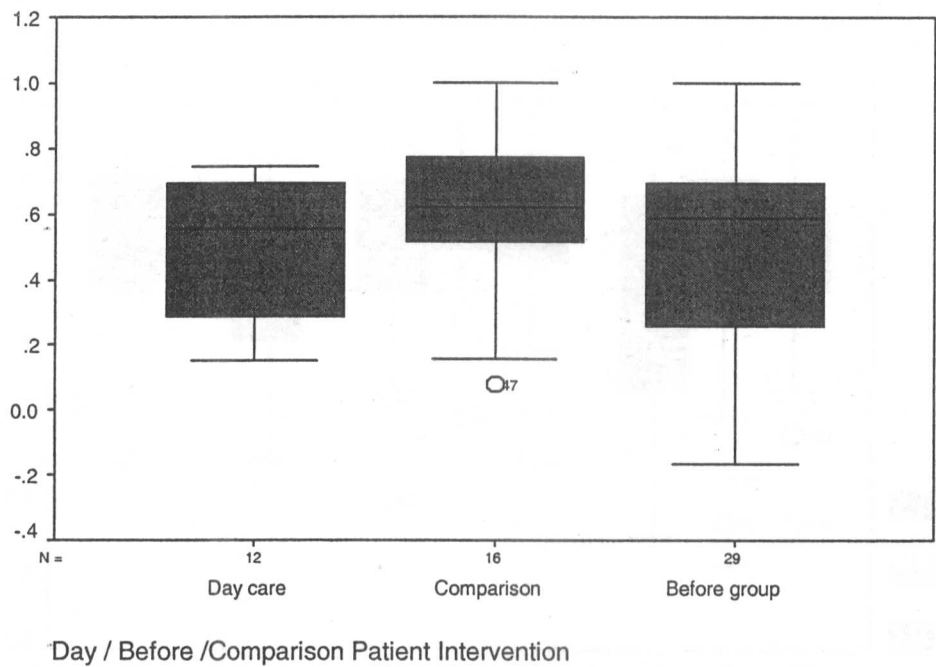
Interview	Intervention group	N=	Mean	Std dev	P value *
Quality of life score at interview 1	PDC group	11	.651	.241	.132
	Before group	25	.506	.293	
Quality of life score at interview 2	PDC group		.696	.154	.188
	Before group		.592	.316	
Quality of life score at interview 3	PDC group		.575	.292	.768
	Before group		.605	.251	

\* equal variances not assumed

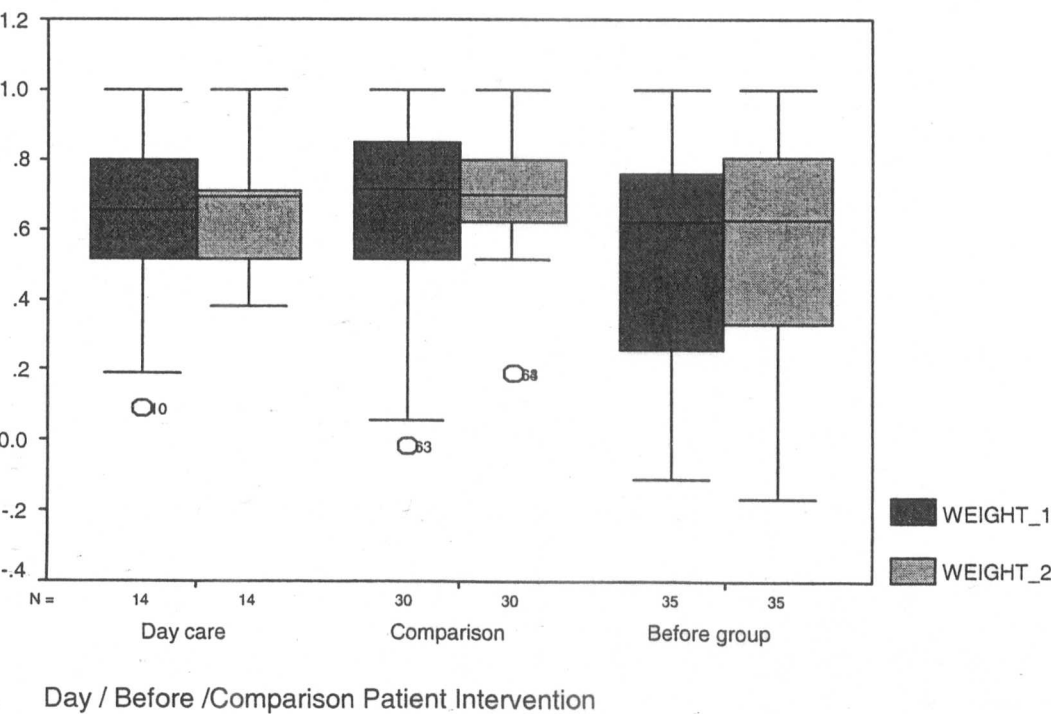
The mean values are slightly lower for the before group, but the results of the t-tests show that, for all comparisons between the PDC group and the before group, none are significantly different. Mean scores are around 0.5 and 0.6 for both groups, indicating that a year in this state is the equivalent of 0.5-0.6 of a year in full health. For patients with more than one interview, the values do not change by more than 0.1 between interviews, indicating that these values are stable and not changing over time.

Since no quantitative difference in quality of life was detected with the EQ-5D instrument, the economic analysis became, *de facto*, a cost-minimisation study. The cost data were not synthesised with outcome data to produce cost-effectiveness ratios.

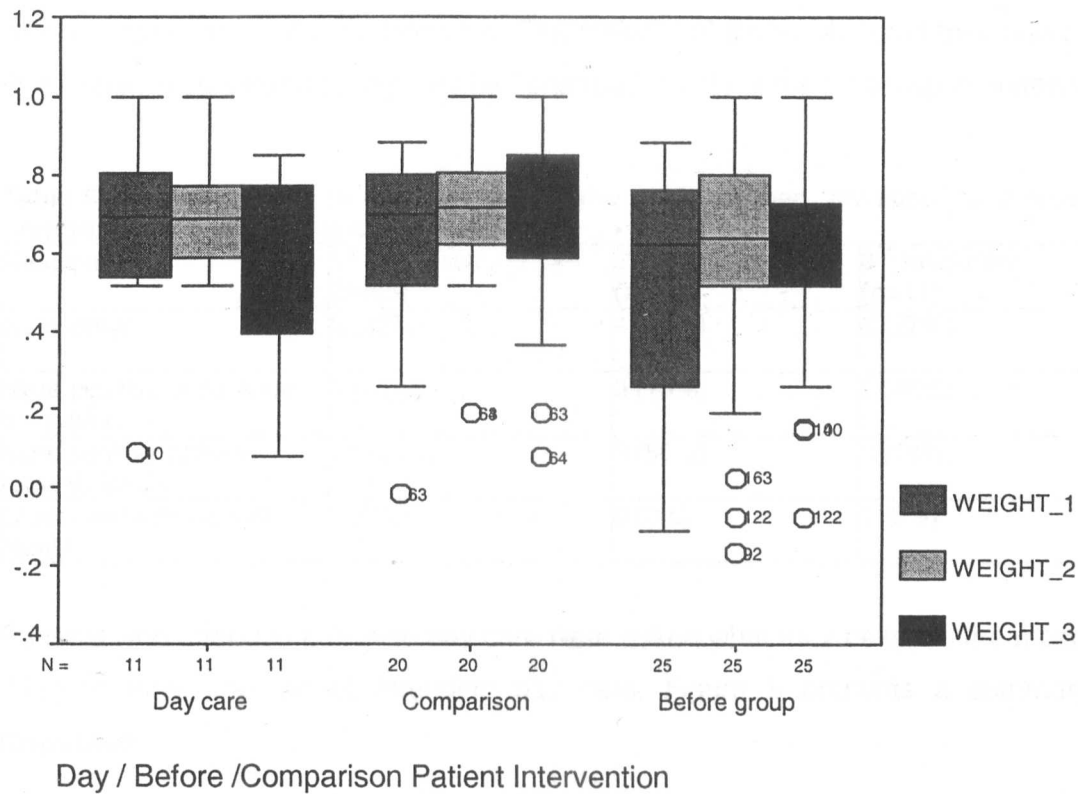
**Figure 3: Box plot of EQ-5D composite scores for patients who have data from the first interview only**



**Figure 4: Box plot of EQ-5D composite scores for patients who have data for at least two interviews (weight 1 and 2 refer to interview 1 and 2)**



**Figure 5: Box plot of EQ-5D composite scores for patients who have data for time points. (weight 1, 2 and 3 refer to interview 1, 2, and 3)**



**SECTION 3. RESULTS FROM THE ANALYSIS OF THE OPEN-ENDED COMMENTS DURING THE INTERVIEW<sup>2</sup>**

At baseline interview, over half of the PDC patients suggested that they had changed as a result of attending PDC. This was described either as a change in outlook (24%) or attitude (44%). Table 6.14 shows the proportion of respondents who said they had changed, by different categories, at each interview. The majority of those who said they had changed at first interview continued to say they had changed (n=11) at the follow-up interviews.

**Table 6.14 Respondents in the PDC group who said they had ‘changed’ as a result of PDC: proportion of responses at each interview<sup>2</sup>**

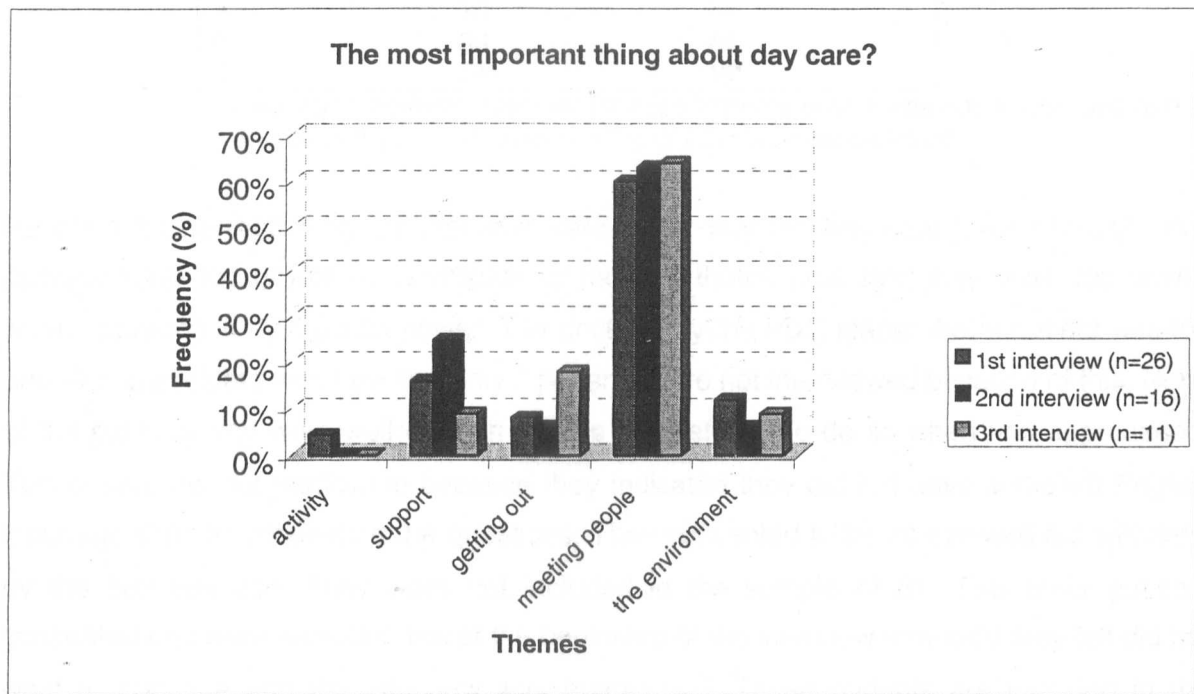
Response	1 <sup>st</sup> interview (n=25)	2 <sup>nd</sup> interview (n=16)	3 <sup>rd</sup> interview (n=11)
No change	8 (32%)	4 (25%)	3 (27%)
More positive outlook for the future	6 (24%)	3 (19%)	5 (46%)
More positive attitude towards illness	11 (44%)	9 (56%)	2 (18%)
Closer relationship with family	0 (0%)	0 (0%)	1 (9%)

Patients who attended palliative day care were asked what they believed the most important thing to them was about attending day care. Figure 6 presents a summary of their responses.

These data show different results from the EQ-5D analysis. It suggests that patients who attend feel they have improved. However the interviews do not provide any analysis of the strength of patients’ preferences for PDC overall or for particular aspects of the service. As such, the analysis is limited. The next section shows results that try to address these issues.

<sup>2</sup> Data collected and analysed by another researcher on the study team, (DG). The results are summarised here for comparative purposes

Figure 6: Patient reports of most important thing: frequency of themes at each interview



### SECTION 4. RESULTS OF THE PDC CHOICE EXPERIMENT

This section presents the results of the choice experiment and interpretation of the findings. The strengths and limitations of this study design generally, and in the specific context of complex services such as PDC are discussed in the next chapter. The process of undertaking the study is reported here as one of the objectives was to evaluate whether this approach would be feasible in the context of PDC. The results of the random effects probit model are reported and interaction effects explored.

#### Sample size

Table 6.15 shows the numbers of patients attending each PDC centre and the number of respondents who agreed to be interviewed. In total, 81 patients agreed to take part across four PDC centres. Centres A and B experienced reduced numbers of referrals during the data collection period. Centre A reduced the numbers of days it opened from four to three, limiting the number of patients who could take part in the study.

**Table 6.15 The proportion of all patients' interviewed for the choice experiment, by centre.**

Centre	Number of interviews	Number of patients attending PDC
A	14	18
B	10	14
C	33	46
D	24	45*

*\* As some patients attended for appointments only it was not always possible to approach them for interview before or after their appointment*

Reasons for not obtaining an interview were either that the day care leader thought that patients were too unwell to participate or patients themselves said they were too unwell when approached by the interviewer. The decision by the PDC leader that a patient was too unwell to participate was rare and only 7 patients were not interviewed because of this. None of the patients who were well enough to take part refused to do so when first approached. Two people did not participate because they indicated they did not have sufficient English language skills to understand the questions. They consented to be interviewed but withdrew by the first question. They were not included in the sample of 81. Two other patients consented and were recruited, but at the beginning of the interview they said they felt did not want to continue and the interview was abandoned. These patients are included in the sample. Altogether 79 patients participated in the choice experiment.

There was some interview fatigue reported by a small number of the more frail patients who consequently did not complete the interview (three respondents). The ordering of the scenarios was altered for centre 3 and 4 to guard against respondents systematically completing fewer of the 'hold-out' choices since they had been choice 9 and 10 for the first 2 centres. One patient (a younger man) chose not to make a decision for half of the pairwise choices, saying that he would not choose either if offered these specific choices. These "choose neither" were dropped from the analysis. All other responses were included in the analysis, whether patients complete all 10 pairwise choices or not.

### **Demographic characteristics**

Of all patients who consented, 46% were male and 54% female. Mean age was 61 years (median 63 years, range 20-89 years). Thirty-eight per cent of patients said they lived alone, 45% with a partner and 14% with other family members. The mean number of months that patients had attended a PDC centre was 8.13 (median 6 months, range 1 month to 9 years). No other demographic characteristics were collected since this data was not included in the analysis.

### The choice models

Table 6.16 shows the results of the random effects probit model. This model includes all patients' choices, regardless of whether they expressed dominant preferences overall. High levels of significance indicate that the presence of an attribute factor was correlated with the respondent's decisions to choose a particular scenario.

The results of the probit model indicate that all the signs on the coefficients are as expected, with the presence of a positive attribute (availability of a doctor, availability of specialist therapies) increasing the likelihood that a respondent would prefer that scenario. The only attribute with a negative sign is "time" which represents a centre's opening hours. It suggests that later opening time of 1pm-6pm (which was coded as 1) is less preferred than 10am-3pm (coded as 0).

The probit model demonstrates that all the attributes except bathing and hairdressing had a significant impact on respondents' propensity to choose scenario A. The most important attribute was specialist therapies (0.6118) which was more than twice as important as staying all day (access: 0.2832) which was more important than routine access to a medical support (Doctor any time: 0.1857). The constant term in the model was not significant, indicating that respondents were not predisposed to choose scenario A or B, which was as expected.

**Table 6.16 The results of the random effects probit model (without the two hold outs): All patients including those manifesting dominant preferences**

Attributes	Coefficient	P value	95% confidence interval	
Access	0.2832	< 0.0001	0.1720,	0.3944
Time	- 0.2987	< 0.0001	-0.4102	- 0.1872
Bath	0.0811	0.154	-0.0303,	0.1926
Specialist therapies	0.6118	< 0.0001	0.5003,	0.7232
Hairdressing	0.0446	0.430	-0.0330,	0.1551
Doctor any time	0.1856	0.001	0.0755,	0.2957
Constant	0.0220	< 0.694	-0.0880,	- 0.5972
No. of observations: 624 (some patients did not answer all questions)				
No. respondents 79				
Mean number responses 7.9 (range 4 to 8)				
McFadden R <sup>2</sup> 0.206				
Log likelihood function -343.38209				
Correct predictions from holdout analysis:				
Y=0	61%			
Y=1	85%			
$\chi^2$	142.03 (p < 0.0001)			
(predicted vs. actual values)				



To assess how long to continue recruiting patients to the study, the probit model was run after the first 25 patients' data had been interviewed. At this point, all attributes were significant, except hairdressing and bathing (with wider confidence intervals). However, at this stage, routine access to a doctor was only borderline significant. Patients from the PDC centre with routine access to a doctor had not yet been sampled at this point in the data analysis. Data from 50 respondents showed the same pattern with narrower confidence intervals. The result did not change after analysis of 75 respondents. It was agreed to stop recruitment at this stage since the same attributes had stayed significant for all three rounds of analysis.

An ordinary probit and fixed effect model were also estimated in STATA. There was no appreciable difference (up to 4 decimal places) in coefficients between the random effects probit model and either the fixed effects model or ordinary probit, suggesting that inter-respondent effects were not important in this dataset. The implications of this are discussed further in the next chapter. The probit and logit models were also run in a different statistical software package (SPSS version 10) and produced the same results (Appendix E).

The coefficients in table 6.16 indicate that access to specialist therapies is the most important attribute in the decision to attend PDC when all patients are grouped together. It is about twice as important as the two next most important attributes which are the timing of PDC (10am-3pm is preferred) and staying for the full session rather than coming for an appointment only.

### **Exploration of effect of excluding dominant preferences**

Case by case analysis examining each choice made by every respondent revealed a high proportion of patients who expressed preference for one dominant attribute, by always choosing the scenario that included this attribute (table 6.17). Overall, nearly half (48%) of all patients interviewed expressed dominant preferences for a particular aspect of PDC. Nearly a quarter of all patients showed a dominant preference for specialist therapies (23%). This is a higher proportion of dominant preferences than has been reported in other choice modelling studies in the health economics literature. This is a result in itself (that PDC patients appear to have strong preferences for one attribute, although the dominant attribute differs between patients) but it may also be an artefact of the design of the experiment, since respondents were only offered two levels for every attribute. This high proportion of dominant responses may have had an impact on the regression model. Therefore a second model was estimated that excluded all those patients who expressed dominant preferences (table 6.18).

**Table 6.17 Dominant preferences observed in the data, all patients**

Attribute	No. patients who <u>always</u> chose the scenario with this attribute	% all patients, i.e. including traders n=79	% patients who expressed dominant preferences only n=38
Access	8	10%	21%
Time	8	10%	21%
Bath	0	0	0
Specialist therapies	18	23%	47%
Hairdressing	0	0	0
Doctor <sup>1</sup>	4	5%	11%
<b>Total dominant preferences</b>	<b>38</b>	<b>48%</b>	<b>100%</b>
Total 'traders'	41	52%	
Total	79	100%	

<sup>1</sup> All the patients with a dominant preference for a doctor were from centre D, one of the two centres offering routine medical appointments

**Table 6.18 The results of the random effects probit model: Only those patients who did not display dominant attributes (n=41, including those with missing choices)**

Attributes	Coefficient	P value	95% confidence interval	
Access	0.2795	<0.0001	0.1234,	0.4357
Time	-0.3247	<0.0001	- 0.4811	- 0.1684
Bath	0.1362	0.088	- 0.0201	0.2926
Specialist therapies	0.5269	<0.0001	0.3705	0.6833
Hairdressing	0.1030	0.191	- 0.0515	0.2574
Doctor any time	0.2316	0.0001	- 0.0768	0.3865
Constant	0.0477	0.545	- 0.1068	0.2022
No. of observations	344			
No. respondents	41			
Mean number responses	7.8 (range 4-8)			
McFadden R <sup>2</sup>	0.1014			
Log likelihood	-174.73517			
Correct predictions (from holdouts):				
Y=0	78%			
Y=1	69%			
$\chi^2$	163.40 ( p< 0.0001)			
(predicted vs. actual values)				

The magnitude of the coefficients is similar in tables 6.16 and 6.18 and the coefficients have the same sign. The coefficients for hairdressing and bathing are not significant in either model. The level of significance of the 'doctor any time' improved from 0.001 to 0.0001. But the significance of the other attributes remained below 0.0001 for both models.

In table 6.18, the confidence intervals are wider (fewer observations to model from) but the significant results and the overall findings do not change. The McFadden psuedo-R<sup>2</sup> statistic is 0.10 in this model whereas it is 0.21 in the model with all patients' responses included

(Table 6.16). This suggests that the second model is a worse fit than the first. The interpretation of this statistic is explored in the next chapter.

The results suggest that staying at a PDC centre for the full session is preferable to coming for an appointment only; 10am-3pm is preferred to a later opening time later in the day. Access to specialist therapies influences patients' decision to attend, as does access to the doctor any time. Hairdressing and bathing do not contribute to the decision to attend PDC. This is a highly plausible result as table 6.19 confirms that these two services are the least used for all patients in the study.

**Table 6.19 Activities accessed in PDC by respondents in the study (N=79)**

Activity	No. respondents	%
Social interaction	60	76%
Doctor	41	52%
Physiotherapy	36	46%
Massage	32	41%
Reflexology	24	30%
Aromatherapy	28	35%
Counseling	16	20%
Arts/crafts	33	42%
Hairdressing	12	15%
Bath	7	9%

This is a different result from the findings of the North Thames PDC study that reported higher proportions of patients accessing these services, especially bathing and hairdressing. This issue is explored further in the discussion.

### **Hold-out scenarios**

A comparison of the prediction of respondents choices for scenario 9 and 10 (that were not used to estimate the probit model) were compared with actual choices for these pairwise choices made by respondents. Chapter 5 demonstrated that the dependent variable Y can be interpreted as a propensity to take one action (in this case, to choose scenario A), and at a certain point in a scale, the decision will be to take the opposite action. The functional form of the model provided values of Y for scenarios 9 and 10. The STATA program specified that if Y is greater than 0, then the model would predict that a respondent would be more likely to pick scenario A. If Y is less than 0, then a respondent would be predicted to choose scenario B.

Table 6.16 and 6.18 shows the percentage of correctly predicted values for the hold-out scenarios. This test is run for all respondents and for traders only. Chi-squared tests indicate the likelihood that this result could have been arrived at by chance. It shows that it was

highly unlikely in both scenarios that these predictions could have been arrived at by chance only.

### Analysis of interactions

Interaction terms were entered into the model to explore whether specified relationships between attributes and specific covariates were significant. Significance was explored at the 10% level since the sample size was small and interactions were not expected to show a strong relationship.

Table 6.20 shows the results of adding interaction terms to the model for all respondents (including those with dominant preferences). The coefficients indicate the positive relationship between patients/ service characteristic and the importance of an attribute. It indicates that only the interaction term that combined the younger age group and preference for specialist therapy (YOUNG THERAPY) was significant at the 10% level.

**Table 6.20 Probit model interaction terms explored in the data, all patients, scenario 1-8 only**

Interaction term	Coefficient	P value	Confidence intervals
OLDACCESS	0.0012	0.993	- 0.2674, 0.2650
YOUNG THERAPY	0.2970	0.008	0.0788, 0.5152
YOUNGACCESS	0.1986	0.074	- 0.0192, 0.4164
ALONEACCESS	0.0777	0.449	- 0.1477, 0.3031
MECENTREDR	0.0936	0.407	- 0.1277, 0.3149
ALLDAYACCESS	0.1143	0.448	- 0.2088, 0.4373

### Importance of individual attributes

The marginal rate of substitution (MRS) is interpreted as the ratio of coefficients in the regression model. Without a continuous variable, the marginal rate of substitution for the probit model that has been estimated in this study is not straightforward to interpret. It would be more accurate to consider the relative importance of the individual attributes. Since the goodness of fit appears to be better in the first model and the model uses data from more respondents, these data were used to estimate the relative importance between specific attributes of PDC. The coefficients in table 6.16 indicate that regular appointments with a doctor in a PDC are a significant attribute in decisions to attend, but is reported as the least important of the four significant attributes. The ratio between attributes indicates that it is about a third as important as specialist therapies when all patients' are included in the analysis (ratio =0.30), but the overall preference for access to a doctor any time appears to

be higher when patients who have dominant preferences are taken out of the analysis (ratio =0.44). The interpretation and usefulness of these statistics will be explored in the next chapter.

## **SUMMARY**

This chapter has reported the empirical results of the EQ-5D and choice experiment studies in the Chichester study. Both parts of this work were successfully completed and there were no practical difficulties in recruiting patients to either study, except that the numbers attending PDC were lower than expected. The North Thames study undertaken before the Chichester study had faced many of the problems of other evaluation studies in palliative care research: poor recruitment, poor follow-up and problems of identifying an appropriate comparison group. The Chichester study was an improvement in a number of respects: first, in 'before-and-after' study, the main comparison group (the 'before' group) was more easily identifiable since the PDC centre did not exist in the first phase of data collection. Second, more time was spent in the PDC centre with the staff who would help recruit patients on defining the process of recruitment and on monitoring participation. Since the quality of life study was focused on one centre only, the process was more straightforward.

The choice experiment study was the first of its kind in palliative care. The use of this methodology in palliative care research has been an experiment in the process as well an attempt to obtain relevant and meaningful insights into the strength of patients' preferences for specific aspects of PDC. Interviews rather than postal questionnaires provided high levels of recruitment and completion interviews and ensured that respondents understood the task.

The next chapter discusses the importance of the findings and addresses the question of validity and the usefulness to these approaches in the context of palliative day care. The final chapter (chapter 8) considers the empirical findings and in light of the theoretical debates and methodological discussions outlined in the earlier chapters of the thesis.

## **Discussion of the empirical findings**

### **Introduction**

This chapter discusses the findings presented in the previous chapter and considers how this knowledge contributes to the understanding of how palliative day care centres should be organised and how the service might be evaluated in future studies. It judges whether the health gain/ quality of life data from the empirical research has provided sufficient evidence to demonstrate that this is or is not a useful way to proceed in evaluation of PDC and similar complex interventions. The discussion then focuses on whether the choice experiment evidence provides any additional or different insights into the value of PDC services that might persuade future researchers or funders of research that this could be a useful way to proceed in economic evaluation studies. The methodological design of the choice experiment study is considered in light of the findings, to assess whether it could have been undertaken differently with hindsight or additional information, especially with regard to the decision not to include a price as an attribute in the questionnaire. The final chapter considers these results in the light of the evaluation of complex services and the lessons that can be drawn from using the health gain and choice experiment approaches.

### **SECTION 1. THE FINDINGS OF THE COST ANALYSIS**

The main problem with the cost analysis was that the sample size was not adequate to be able to make any judgement some about the likely incremental cost (or savings) associated with attending PDC. The challenge of estimating an appropriate sample size when undertaking costing studies within effectiveness studies has been highlighted in the health economics literature, largely in connection with large clinical trials (Al et al 1998, Willan 1999, Willan 2002). The particular problem faced here is that a larger sample size was not feasible within the resource limitations of the study. What may be concluded here is that no evidence has been produced that suggests there is a large difference in costs between people who attend PDC and people who do not. Sample size in studies that include an economic component should be designed in such a way that they are powered to detect differences in costs. But the issues associated with sample size calculations for cost-effectiveness analysis have not received the same attention until recently as clinical evaluation (Briggs and Gray 1998). Briggs and colleagues have explored a formula for deriving sample size from confidence interval limits on costs and effects in a trial (Briggs et

al 2002), but it difficult to see how *ex ante* these could have been estimated for this study. Such a calculation would nonetheless have strengthened the findings of the cost analysis.

Therefore, the economic analysis focuses on the evidence of the effectiveness or value of the service, given that the incremental cost of care (a separate issue from the funding of care) for people who attend PDC and those who do not, does not appear to be very large for this sample of patients.

## **THE FINDINGS OF THE EQ-5D STUDY**

### **No significant differences in outcome using EQ-5D**

The analysis of health outcomes using EQ-5D did not show significant differences in health outcomes between palliative care groups. P-values of the main quality of life scores were not significant at the 10% level (a relatively low threshold for significance) for any of the sub-groups of patients who survived and participated in the study at baseline, or in the two following interviews.

The estimation of a cost-effectiveness ratio has not been achieved for this study because the cost results and the effectiveness results failed to show any differences between groups. The EQ-5D findings however, cannot be seen as conclusive evidence that this is an inappropriate methodology as the numbers of patients recruited to the study was small, especially when patients were sub-divided by number of interviews as a proxy for proximity to death. The sample size was calculated for the EQ-5D study *post hoc*, based on the actual EQ-5D data obtained in the study at the second interview (the number of respondents at third interview was reduced due to death or withdrawal from the study). To have been able to detect a difference between the 'before' group and 'palliative day care' groups, using the actual mean values and standard deviations in the study, at a 10% level of significance, the sample size would have to have been 139 respondents in each group, interviewed at least twice. Given the fact that the rate of withdrawal from the study by time two was 140 respondents at baseline to 82 respondents at second interview, the proportional number that would have to be recruited to ensure having a sample of 139 at second interview would be almost 240 in each group. Given the problems of recruitment in palliative care studies described in earlier on, this was not a feasible or practical sample size for a palliative care study. It raises the question of whether the EQ-5D and other global measures of health related quality of life that require larger sample sizes than can be achieved to detect what is a very small difference in health status are useful tools for areas of research which is characterised by small scale studies.

One positive finding concerning the EQ-5D was that there was almost 100% completion of the interview schedule by participants in all palliative care settings and for all interviews. This means that the lack of significant differences between groups was not an artefact of differences in questionnaire completion rates.

One approach to obtaining more robust evidence might have been to group all patients together to increase the power of the study, and to analyse each health domain separately for significant differences between groups. But the reason for not taking this approach can be illustrated by table 6.8 and figures 4 and 5 in the previous chapter. Table 6.8 shows the decomposed results of the EQ-5D questionnaire and the percentage of patients reporting themselves to be at each level of quality of life in the EQ-5D scoring system. In the domain 'usual activities', the PDC group appears to have a dramatic improvement between interview 1 and 2: the proportion of patients facing extreme difficulties in carrying out usual activities is 23.1%, which falls to 6.3% by the second interview. This could well be found to be significant if statistical analysis were applied to these data. However, the sub-group analysis shows no such commensurate fall in *overall* quality of life weights (figures 4 and 5). The large drop in scores could be a result of patients who experience the most acute problems dropping out of the study (due to becoming more ill or dying) and not being interviewed after baseline.

Furthermore, the data were analysed in sub-groups because this made more sense in a context where patients at different phases of illness (proximity to death/ acute illness, as measured by the number of completed interviews) may have a different intensity and range of services (in and out of PDC), and the effectiveness of PDC might be different for different groups.

The capacity of the EQ-5D study to detect differences between palliative care groups was also compromised by the fact that the study was not powered for the EQ-5D study. This means that the negative finding (of no significant difference between palliative care settings) was compromised. This issue is returned to in the final chapter in considering the limitations of the study.

The correlations with the Palliative Care Outcome (POS) scale showed that, for domains that were comparable across instruments (pain and anxiety), the data were strongly highly correlated for pain at baseline and second interview but by third interview, this correlation was not significant at the 5% level, which may be due to the small numbers left in the study. For anxiety, the pattern was not as clear as some data showed evidence of correlation but



some did not. Without prior hypothesis about which groups would or would not be expected to have correlation at each interview, this result is not easy to interpret.

There are other correlations that could have been explored in the data. Living alone could have an impact on the intra-group variation (measured by the standard deviations around the mean composite EQ-5D scores). Also, within the palliative day care group; the intensity of PDC could be correlated against the EQ-5D scores to explore whether changes between scores over time in the PDC group among those patients who access more activities services (medical, social, psychological services) within PDC. The problem with this approach is the method of measuring intensity: PDC is set up to allow people to try out different activities, some of which were by appointment, some of which could be freely accessed at any time. It would have been useful to measure the number of hours attendance per week for each individual attending day care as a possible proxy for service intensity. Respondents' age and state/stage of illness and proximity to death could also explain the variation in scores.

### **Were the findings more robust than in the previous North Thames PDC study?**

It is argued that the study design represents a more robust research design than the first multi-centre study of palliative day care that had no 'before' group. Consequently, the fact that the results still show no measurable effect differences could be considered to have more weight. Furthermore, the use of a global health-related quality of life measurement in this study might be expected to reveal a more general improvement than disease-specific instruments that can only present decompositional scores. EQ-5D can measure more general dimensions of health and welfare that could reflect the aims of palliative day care more than the disease specific instruments.

### **Qualitative versus quantitative findings**

There are two reasons why the quantitative quality of life results data may not have shown a difference between intervention groups: first because palliative day care does not actually produce improvements in health related quality of life. Second, the important dimensions of quality of life may not be measured using this methodology. If the first explanation is true, this does not mean that palliative day care has no value, but that it may be performing a different service and meeting different needs to those originally envisaged.

The open-ended responses by patients showed that palliative day care might be having its greatest effect on the social domains of quality of life. Meeting others was cited by around 60% of all attenders at each interview as the most important aspect of palliative day care.

Patients who attended also reported a more positive attitude towards their illness, and a more positive outlook. None of these domains of psychosocial health are emphasised in the EQ-5D instrument. The designers of the EQ-5D instrument focused on more obvious differences in quality of life, and palliative day care is not the setting in which the instrument works at its best.

The lack of evidence of difference in quality of life between patients who attend PDC and those who did not contradicts the qualitative evidence from the open-ended statements collected during the same interview. This supports the argument that EQ-5D may not be an appropriate way of measuring the outcomes of PDC. Since changes in quality of life from PDC might be subtle, the sample size in PDC evaluation studies should be larger than for studies of interventions that are expected to have larger effects. However, given the number of patients who attend PDC on each day (and that centres are open 3-4 days a week), a multi-centre trial would be required to recruit an adequate sample size. The value of undertaking multi-centre rather than single centre studies may be an important consideration for future evaluative studies of PDC.

### **Limitations of the EQ-5D study**

The Chichester study was a robust study design but the numbers of subjects in the survival sub-groups was small: between 11 and 36 in each group. This is a possible source of bias in these findings. Those who are more ill were more likely not to have been interviewed after the baseline interview and only those whose health did not deteriorate rapidly were represented in the sample of patients' interviewed at the second and third time points. As the sub-groups with two and three interviews represented patients who were relatively well, it would be expected that there would not be any large differences in health-related quality of life between PDC attenders and non-attenders. It might have been a different pattern if patients whose health deteriorated rapidly could have been interviewed, but there would be serious practical and ethical problems in conducting interviews in these circumstances.

The study was not powered to detect differences in outcome using the EQ-5D instrument. When a range of outcome measures is used and no specific clinical outcome is seen as dominant, the choice of sample size is a problem. The EQ-5D study was seen as an 'add-on' experimental addition to the quality of life study design that was intended to mirror the original North Thames study. Also, since the Chichester study was undertaken in one health district, the main consideration was being able to maximise the numbers of patients who could be entered into the study. The original target was 210, 70 in each group. But considering the number of new referrals to day care (approximately 10 per month), it was not

possible to recruit 70 patients to the day care group without recruiting for at least 18 months and this was neither a realistic nor a cost-effective option for this study.

Since health gain/quality of life approaches to evaluating palliative care will continue to be the dominant methodology for evaluating effectiveness, and since the EQ-5D instrument is easy to administer, it would be worthwhile testing the instrument on a larger sample size of patients, for little additional cost or effort. This would be a valuable contribution since this PDC study is the only study of this kind in palliative care research to adopt this approach to measuring outcomes.

## **SECTION 2. THE FINDINGS OF THE CHOICE EXPERIMENT STUDY**

### **The attributes of PDC that were seen as relatively more important**

The results of the study suggest that patients had strong, sometimes dominant preferences for particular characteristics of PDC. Specialist therapies (aromatherapy, reflexology and massage) were seen to be the most important service characteristics for all patients, and especially for younger patients. Other characteristics such as living circumstances and the type of centre respondents attend did not appear to have significant association with preferences for care.

Overall, a regular appointment with a doctor was perceived to be less important to patients than specialist therapies. All the patients who expressed dominant preference for access to a doctor were from the centre that offered regular medical appointments with a palliative care consultant. Staying all day was a positive attribute of PDC, and opening hours of 10am to 3pm are preferred to later in the day. This is the first quantification of the strength of preference of these attributes and this should inform providers about which services they provide should be prioritised. However, additional information on cost-effectiveness and other criteria (for example the views of other people who might benefit from PDC, such as carers) would also be important information to inform decision-making.

### **Characteristics that were seen as relatively less important**

One of the issues raised by providers of the service (and hence the inclusion of this attribute in the experiment) was whether the PDC centre's opening hours were best suited to the convenience of the staff (who also had to undertake inpatient ward rounds early in the morning and in the later afternoon, as well as administrative tasks), rather than around the needs of the users of the service. But the evidence from this study would suggest that these daytime hours were preferred by patients to opening hours later in the day. This preference could have been different if the views of carers had been included in the study. Carers might

have preferences for either longer opening hours or times that are more convenient for them to do paid work. If PDC was also offered as a way to support home carers, this might be explored in a future study.

The results showed that bathing and hairdressing were not as important attributes as other aspects of day care according to patients. This was not surprising as both these services can be accessed through other means (privately or through social services). These attributes were added to the experiment as it was not clear whether these personal care services were important to patients and whether they should be provided in PDC centres. Some centres did provide them for almost everyone at one point in time, and some did not. At one centre where hairdressing had been withdrawn due to the lack of a volunteer hairdresser, patients had commented to researchers that they would be willing to pay for a hairdresser privately in order for the service to continue. The strength of demand for this service overall was not known before this study was undertaken.

The data did not indicate that respondents who attended centres with routine medical appointments were, overall, more likely to prefer medical appointments. However, the only respondents who displayed dominant preferences for routine appointments with a doctor were in the centre where this was available, but this was only 4 out of 24 patients who attended that centre. Similarly, attendance at a centre where patients were encouraged to stay all day did not appear to impact on preferences for this attribute. This would indicate that the philosophy of care of the centre patients attended did not bias their demand for particular attributes of care.

### **Contribution and additional insights gained from the choice experiment**

The choice experiment has demonstrated that the quantification of patients' different patterns of preferences and the systematic comparison of different aspects of the service revealed a different pattern of demand for services or attributes than simply asking people about their preferences. When faced with constrained choices, respondents showed that personal services (bathing and hairdressing) were the least important to their decision to attend. This leads to a different outcome from one that would have been arrived at by simply listening to respondents' comments in interviews or casual discussion, where they were not being asked to make choices or trade between services or attributes. For a service such as PDC that is heterogeneous and where each centre has its own philosophy and has strong proponents for particular aspects of care, this is an important finding. A future study might consider running an empirical investigation to establish whether open-ended questionnaires

about patients' preferences for different aspects of PDC arrive at a different conclusion to those found from a quantitative choice experiment.

### **The influence of experience of different services on demand**

One of the characteristics of PDC was that the range of provision of activities and services changed over time and patients might not have had experience of all the attributes presented in the choice experiment. The therapies and services offered at each centre were determined to some extent by the supply of local therapists with particular skills (paid for their time or working as volunteers), and the philosophy of the PDC centre, rather than demand for these particular services. For example, at the start of the North Thames PDC study, bathing and personal care were common in the five centres that participated in that study and the uptake was high in some centres with virtually all patients having a bath during their visit. In the CE study (undertaken three years later), none of the participating centres were offering baths on a regular basis. Since the service was not offered, there is an argument that attenders would not have experienced the positive or negative consequences of this service and therefore may not have known the value of it compared with other services they did know and use. It would have been insightful to be able to conduct the study by including patients who had experienced PDC with bathing and hairdressing offered in the recent past, and then interviewed them at a time when these services were no longer offered. This could have established whether their preferences were different from those who had never experienced these activities since these were the attributes that did not affect choice of scenario. One of the advantages of the choice experiment technique is that it can estimate the value of services that have not been experienced if the value of the attributes of a service is relatively straightforward to evaluate by individuals. It would have been insightful in this context to explore this further by assessing whether the value of bathing and hairdressing was different for those who did experience them in the context of the PDC environment.

### **The robustness of the findings**

The choice experiment results indicated that specialist therapies were the most preferred or valued aspect of palliative day care. This was true when respondents with dominant preferences were included in the analysis, and when they were excluded. None of the results altered when the dominant preferences were removed. Since a smaller dataset was analyzed for the model that excluded dominant preferences it would be expected that the confidence intervals around the parameter estimates for each attribute would be wider. This was found to be the case.

The only slightly different result between the models was the “doctor any time” attribute that became even more significant in the second model (without dominant preferences) but this is not straightforward to interpret. What it means on face value is that, for patients who were willing to trade between attributes, the presence of a doctor any time was more highly valued, and contributed more to the decision to attend the centre with this attribute present, than when dominant preferences were included in the analysis, but it still remained highly sensitive.

The usefulness of the constant term in a model without a constant status quo scenario (for which respondents may have a natural preference) is debatable. The STATA model produces a probit model with the constant term as a default in the program. Its usefulness in this context is only to show that the constant term is not significant, indicating that respondents did not exhibit so called ‘irrational’ choices by always choosing A or always choosing B, regardless of the attributes presented in them. Since latent utility is not observable, the analyst cannot estimate the utility function directly and can only represent the deterministic part. The constant variable element acts as a check on whether the model is mis-specified.

The usefulness of the marginal rate of substitution (MRS) is that it can indicate the relative strength of preference for one attribute over another. In this context, the attributes represented descriptive characteristics of the service. They were either present or absent. In this example, the substitutability of attributes is limited and the meaning of the MRS is also limited. The usefulness is in the interpretation of the strength of preference of respondents for attributes in the face of constrained choice. It is an improvement on rating exercises as not all attributes can be rated as equally important. It is also an improvement on ranking exercises as it provides some indication of how much more one attribute is preferred to another. In the context of this choice experiment, without a price attribute, the MRS cannot be interpreted beyond this. The results show that specialist therapies are strongly preferred, and that the strength of preference is between two and three time stronger for this attribute (access all day, the opening time, and the presence of a doctor every visit). This should provide a signal to decision-makers that this is the attribute which patients value highest, and therefore should be offered routinely if the service aims to provide services that are valued by those who use them.

### **The appropriateness of the methodology**

The purpose of the choice modeling experiment was to begin to assess systematically the value of particular components of PDC by considering patients’ preferences for these

attributes. The research question was not framed in such a way as to determine whether PDC should be offered to all patients (its overall effectiveness) but how it could be organised to reflect patients' expressed preferences for particular attributes of the service that providers could influence. Measurement of these preferences may be the closest approximation of the measurement of benefits where objective measures are not readily conceivable or meaningful and where giving people what they want may be a more important service than giving them what is effective or 'good' for them, as determined by others. So long as the aspects of care that are valued by patients who attend as contributing to utility (by whatever means), it is hard to argue that these services should not be provided to people near the end of their lives. Since these services have traditionally been provided outside the statutory health sector, by voluntary organisations, this is evidence that services for people at the end of life are also valued by people who are not experiencing this state of illness themselves.

Providers see aspects of the provision of palliative day care such as its contribution to existential health or to aspects of psychosocial well-being as important objectives of palliative day care. These attributes are also contested by different stakeholders in the service, some of whom believe may believe in them strongly, some of whom are more sceptical about the contribution of palliative day care to these goals. The extent to which, for example, public funds should be used to provide spiritual well-being and contribute to wider social goals related to quality of life is not straightforward. Whether or not PDC services are valued by those who attend for these specific attributes (in terms of whether and how much people might trade in order to have these benefits) is not known. The challenge for empirical research would be to capture these subjective and highly context specific attributes of care in a choice experiment that is, by its nature, a simplification of the real choice that users face when deciding whether to attend a centre. Health-related quality of life instruments have been developed that measure some aspects of existential health, but no simple, overall form of words that would capture this domain in one attribute has yet been validated. This may be an important research topic in the future.

Clearly, a model without a willingness-to-pay component (or some other continuous measurement of cost, such as waiting time or traveling time) will not yield absolute estimates of utility. In this study, there was no continuous variable attribute such as a financial charge or travel time that could be used to estimate marginal rates of substitution with other attributes and therefore an absolute measure of preference. This aspect of study design was considered at length throughout the conception, design and analysis of the CE study. It was decided that willingness-to-pay (or some proxy for the sacrificed of wealth) could not be used

in the context of a PDC evaluation. The argument has been presented in earlier chapters that willingness-to-pay is a disputed approach to estimating utility. There are convincing arguments why this approach would not work in PDC due to the context since PDC is a service provided free of charge to patients, the kinds of people who attend who are mainly elderly and retired, and the way in which the WTP attribute might be interpreted by them as a cost of provision rather than a measure of its value.

Before the study commenced, it was not clear whether a monetary attribute would be such an important disincentive to attending PDC that the choice experiment would fail overall. There was some concern that the inclusion of a daily charge as an attribute in the choice scenarios would lead to patients refusing to participate in the survey. Furthermore, it clinicians and managers involved in the study argued that this approach could be inequitable and unethical even in a research context, as it might increase anxiety among patients about future financial charges for PDC. It was not clear whether the choice experiment format would work at all amongst this group of patients, and other research questions could be answered without including a monetary attribute. Therefore, it was decided not to include it in this study. Having completed the study, this view has not changed and it remains a contested issue in the economic evaluation literature.

The number of dominant preferences expressed by patients in the study may be a result of the study design that only included two levels for each attribute. A high proportion of dominant responses have also been reported in other CE studies in health care. (San Miguel et al 1997, Bryan et al 1998, McIntosh and Ryan 1999, Ratcliffe et al 2002). Future studies might consider including more levels of attributes in order to provide levels of attributes rather than all or nothing choices.

The sample size was small compared with other recent choice modeling studies in the health economics literature. Studies with 70-90 respondents with a high degree of comparative validity have been published, although these have all been non-economic (and not random utility based) studies (Rosko, Walker et al 1983; Wigton, Hoellerich et al 1986; Meister, Lausberg et al 2001). In the first round of analysis of the study, the same attributes were significant for 25 patients as for 50 and 75 patients. For 79 patients, the p-values were highly significant (at the 1% level), and remained highly insignificant for bath and hairdressing. While higher numbers of patients would have improved the predictability of the model, especially for the analysis of interaction terms, it does not seem likely that the overall results (and message to decision-makers) would have changed. In this health care setting, it is a reasonable question to ask whether it is ethical to continue to recruit patients to a study



when their responses would be unlikely to change the findings and consequent decisions about how services ought to be organised.

### **Goodness of fit of the model**

The McFadden R-square statistic has not been routinely reported in other empirical studies using choice experiments. Studies that have reported this statistic have found values of 0.4 (Gerard 2002), and 0.14 (Van der Pol and Cairns 1998), so findings reported here are at the lower end of this range. However, as argued earlier, this statistic is best used to compare different specifications of the same model. It suggests that the model that includes all patients including those exhibiting what might be dominant preferences is the better fit. It is not possible to make broad conclusions from this observation since the Pseudo- $R^2$  statistic has to be interpreted with caution in probit models, and there may be too few data from which to make any assertions. But it makes the case for the inclusion of all patients since this has not made a large difference to the overall outcomes, and may have improved the estimation of the relative importance of the important attributes in the model.

The other indication of goodness of fit used in this experiment was the use of two hold-out clauses to test whether the model would predict the number of respondents who chose scenario A or B for each pairwise choice. The chi-squared test indicated that the model predicted correct choices for the hold-out choices and that these predictions were significantly better than would have been predicted by chance alone in both models. The percentage of correctly predicted choices is now a common way of presenting the goodness of fit of choice model data, but estimating this figure using hold-out choices that are not included in the initial analysis to generate the model is not as common. It could be argued that to not include respondent data from two pairwise choices in generating the regression model is a waste of primary data. But the lack of other straightforward methods for checking the predictive property of the model makes this a more attractive approach than would be the case if ordinary linear regression models could be estimated and a more robust goodness of fit analyses applied. It would be interesting to be able to compare these predictions with those generated from other studies if this approach is adopted more widely.

One of the limitations of using a hold-out approach to testing the validity of the model is that the hold-out scenarios have the same structure and format as the main experiment, and the data is collected in the same way. A more robust test might be to collect data from respondents in a different format, that is, using a different approach to gather information on preferences, such as a ranking or rating approach. Furthermore, collecting evidence of

validity using data from the same patients is not as robust as collecting this data from another group of respondents for comparison.

External validity cannot be tested using a hold-out approach. External validity examines whether the results of the experiment can predict what choices patients would actually make faced with the same choices in the real world. One approach for this might have been to consider the services that a sample of respondents actually choose in their package of palliative day care, and to compare this data with the results of the choice experiment.

### **Insights from the empirical analysis**

Overall, the study has demonstrated that the method is workable with this group of patients despite the pilot study on public health students warning that there may be barriers of cognitive understanding to undertaking an evaluation of this kind. The main study indicated that PDC patients who are mostly elderly with a life-limiting illness are both willing and able to undertake a discrete choice task. The interview was quick (around ten minutes) so it did not make the respondents miss any of their chosen activities and was written in simple English. This is a particularly important attribute when collecting data from frail people, many of whom are elderly. Furthermore, the method of data collection did not require patients to be followed up over time, which is an important attribute of research in palliative care. Future use of this methodology is promising.

The results have indicated that specialist therapies were more important than any other activity in day care. The fact that specialist activities are dependent at least in part on volunteer specialists giving their time free of charge and therefore may not be available routinely (whereas doctors are paid) would suggest that day care managers might reassess the status of specialist therapists and consider having them as core service providers in palliative day care centres.

An approach that can take into account patients' preferences for particular aspects of care may be a more appropriate methodology to explore in this context. The choice experiment approach is a way of exploring these questions in a systematic, quantitative way. The study that was undertaken had to be limited in its scope since it was not clear that patients would be able to answer these types of questions, and there was evidence from the pilot study and from providers that these questions would be difficult for patients to answer. However, the evidence showed that the majority of respondents did not have difficulties in answering these questions. To the contrary, this approach was as short as or shorter than the quality of life

questionnaires, and it did not rely on asking patients to participate over a number of interviews, which was an important attribute in the PDC setting.

Faced with almost no other useful findings from previous attempts to use economic evaluation methods, this method provided more insights than had been obtained by methods. The study was not a full economic evaluation, but it provides useful information on the value of attributes of PDC where no other quantitative method has succeeded. For interventions that are multi-dimensional, and where the benefits are not known, this is an important step towards understanding the value of these services to patients. There is a need for more evidence about whether the difference in study design is important and whether they change (or invalidate) the interpretation of study findings.

Future studies need to be based on up to date stated preference choice experiment theory, which is still being developed and tested in the field. However, rigorous designs may have to be traded off against judgments about their applicability in the context of the study setting. The inclusion of WTP as an attribute for this constituency of patients has not been empirically tested. The next phase may be to consider in more detail whether this approach would be meaningful or workable in the context of PDC. Since this first study has been successful, some of the arguments about the feasibility of this study design for this patients group has been answered, so it may be more acceptable to those working in the field to undertake these kinds of experiments. Data collection was quick and simple, follow-up not necessary and the interpretation of the findings relatively straightforward to understand. It may not be possible to consider whether a proxy for overall value (monetary, time, or some other form of sacrifice) might be achievable in this setting.

### **SECTION 3. HOW THE CHOICE EXPERIMENT STUDY MIGHT HAVE CHANGED WITH HINDSIGHT**

One of the questions is whether it would have been possible to consider more attributes and more complex attributes in the study design. Respondents said they experienced very few problems in answering the pairwise choice questionnaire presented to them. It would have been informative to explore how many pairwise choices they could have answered before they became fatigued. A future study could do this. It would have to be planned carefully, since it would not be ethical to deliberately continue to ask questions until patients became fatigued because it might cause potential harm. But it might have been appropriate to schedule a break after answering the first ten questions and then ask respondents if they would be prepared to answer a few more questions. This could be repeated after another break, still emphasising to respondents that they should only continue agree to continue if

they wanted to and of their own accord. It could provide useful data on the study process for other research in this area. If patients could have answered more pairwise choices without additional fatigue, this would have provided more of an opportunity to test the validity of the study, or to test the axioms of rationality. Since research is now underway in other choice experiment studies to explore whether these axioms are violated in choice experiments, this analysis would have contributed to this knowledge.

The open-ended comments gathered as part of the study has suggested that patients attend PDC for specific services but also because of the social environment and because it provides a way of meeting people in similar circumstances. One of the developments that could be considered in a future study is whether these less concrete aspects of care could be incorporated into the study design. This would also mean having to describe these attributes in ways that were not open to differences of interpretation, and where different levels of attribute could be described in meaningful ways. It has already been discussed that respondents would have some difficulty choosing a scenario in a pairwise choice that did not have a friendly and social atmosphere. Intuitively, it would not seem plausible for someone to choose to attend a centre that was not friendly and social. It might however be possible to describe characteristics that contribute to this attribute. If the demand for a social atmosphere could be broken down into these more concrete constituent parts this might make it possible to consider the strength of preferences for these abstract attributes with more clarity.

It may be possible to undertake more in-depth qualitative interviews or focus groups among attenders of PDC and among prospective attenders to establish if there are any attributes of PDC that reflect or represent these attributes. Qualitative research in economic evaluation is an emerging field of enquiry and may be a valuable area of synthesis between disciplines, especially around the meaning of outcomes, values and preferences (Coast 1999).

A future study could undertake an in-depth analysis of the reasons for people attending to explore their stated reasons in some depth and consider how to dissect the concepts of attributes such as social environment and friendly atmosphere into constituent attributes that affect patients' lives and consider what the meaning of these broad categories in patients' lives. For example, people may choose to attend a PDC centre with a social atmosphere for different reasons. Patients who live alone may have acute needs for social contact and human comfort. Patients who are cared for by relatives may be considering their needs as well as their own.

A 'thinking aloud' study similar to that designed by Ryan and colleagues (Ryan 2001) could ask patients to say out loud what their thinking processes were while they were deciding which scenario they would choose, and to explain the reason for their choices. It would be quite straightforward to record interviews in the PDC study and this may have been an efficient way of exploring rationality of preferences and stability of preferences qualitatively.

Another area of interest that has emerged from this study is to explore the values and preferences of people who attend once and do not attend again. Health and social care professionals whose role it is to introduce the idea of attending PDC to patients (the home care nurses predominantly) have reported that some patients refuse outright to attend, some consider the idea for weeks before attending, and some attend once and do not return. The preferences of individuals who never attend or only attend once have not been explored here or anywhere else. For them, there may be strong disincentives to attending PDC. It would be highly informative to expand the study to include these patients. Instead of describing PDC directly, which they would not be familiar with, the study could present the attributes without describing them as explicitly part of a PDC service. This might find out what configurations of service the non-attenders might respond to more favourably if they were to access specialist services outside their home.

It would also have been informative to ask respondents to rank or to rate the attributes before they undertook the choice experiment questionnaire. This would have indicated how similar or different constrained choice and unconstrained (ranking or rating) choices were for particular attributes. It could have illustrated the added value of a method that forces respondents to think in terms of trading off different aspects of their care. This would have indicated that the approaches measure different constructs, give different kinds of information, and demonstrate why this more complex approach to assessing preferences provides additional information. To interested parties (such as PDC managers and sceptical researchers) for whom the concept and measurement of strength of preferences rather than health gain is not an obvious choice for evaluation, this could have provided important additional information to illustrate the benefits of using of this approach

One of the limitations of the study design was the lack of a mechanism to estimate the monetary value of the benefits provided by palliative day care. One of the aims of the study was to assess whether it would be possible to identify an appropriate way of representing this attribute and developing a cost-benefit analysis approach to evaluating PDC. It would have been enlightening to ask patients directly if they would have accepted a small charge for PDC in order to inform this aspect of the study but this was not possible. There is

anecdotal evidence from undertaking the study that those who support PDC believe strongly that the service should be provided free of charge and that it would be harmful to ask patients questions asking them to consider a daily charge or some other form of payment. It would have been good to establish the strength of feeling from respondents about this aspect of service provision to have more robust evidence of this.

It is still important to try and identify some measure of some representation of sacrifice, financial or otherwise, in order to be able to begin to quantify the value of these services in ways that are comparable with the use of these resources in other contexts. The value of palliative day care compared with other forms of palliative care, cancer care or other sectors, has not been demonstrated in this thesis. It is useful to reflect on what possible proxies for financial sacrifice might be in the context of palliative day care.

It was argued earlier in the thesis that travel time had a low opportunity cost in the context of PDC. A longer journey may even be a positive experience rather than a sacrifice in time (of which there is relative abundance on a day to day basis). However, there would be a point at which a journey time would entail sacrifice (if it became uncomfortable), and it might be possible to identify the levels of journey time that would mean that people would be prepared to sacrifice other attributes of palliative day care in order to avoid such a journey. Other ways of conceptualising sacrifice might be in terms of other resources that are accessed by people who attend PDC. For example, there might be trade-offs between home visits by a specialist nurse or GP and specific attributes of palliative day care (say, specialist medical consultation in the centre by appointment). People might sacrifice other forms of social care for access to palliative day care services. This would not provide a willingness to pay estimate per se. However, there may be innovative ways of addressing this. The idea of giving respondents an amount of credit to spend on a package of health and social care (and even, more radically, that they could theoretically cash in for money to spend outside the public sector on private goods) may gain credibility in evaluative research the future. A choice experiment could include qualitative attributes such as specific services, and an attribute representing the total price of that package of services. This price would not be paid for by the individual out of his or her own pocket, but would present a proportion of their health and social care credit, which theoretically would be the same for all respondents with the same level of need. It might be a way of addressing the problem of ability to pay in deriving marginal benefits in monetary units.

PDC may also have an important role in giving carers' time off and this aspect of the value of the service has not been considered in this study. Other people who may derive value from

the service are the volunteers, people who give altruistically to charitable hospices, and the staff. This relates to the question of whether the provision of a service such as PDC has a moral as well as a quality of life dimension or purpose. Altruism, self-sacrifice, the belief in doing good for others are ideas that are embedded in the religious/ humanist culture of the hospice and palliative care movement (Rees 1982, Abel 1986, Seale 1991, Jennings 1997). Yet these values cannot be considered in the usual paradigm of economic evaluation and, as such, studies that attempt to conceptualise and measure the value of these services may be missing the important aspects of value that these services provide.

The empirical studies have focussed on the outcomes from the users' point of view and a different approach would be to attempt to establish what society would be prepared to pay for these services. Since PDC is usually provided and at least partially funded through the voluntary sector this is evidence that society values the care that people with advanced cancer receive since they make charitable contributions towards it. Whether they have preferences at the micro-level of decision-making, that is, how the contribution is spent, is questionable and has not been investigated. Since the general public would not usually have experience of PDC, they might not be expected to know the value of specific attributes of the service. Furthermore, they might have preferences for the users themselves to have what they feel that they need at that time of life, which brings the question back to the preferences of the patients themselves.

Other considerations are also important: first, whether the users of the service are the most appropriate group from whom to obtain the valuation of PDC. There is a clear argument that since it is the social valuation of services that needs to be determined, these values should be derived from "citizens" rather than patients (Mooney 1997). The question remains about whether the general public would have complete preferences and would be able to make informed choices in such an experiment. PDC might be especially vulnerable using this approach, since the general public may not know such a service even exists, let alone current and future users might value it.

#### **SECTION 4. INSIGHTS FROM THE EMPIRICAL ANALYSIS FOR OTHER HEALTH AND SOCIAL CARE SECTORS**

The thesis set out to consider the challenges of economic evaluation of PDC as an example of a complex service. As such it is useful to reflect here on what lessons might apply to the evaluation of other complex areas of health care. First, the EQ-5D instrument may not be an appropriate tool for measuring outcomes for interventions where large changes in quality of life are not expected. As each dimension has only three possible levels, it may not be

sensitive enough to detect important changes in quality of life for interventions where patients/ clients are expected to continue to experience “some problems” (and to tick this box in the questionnaire).

The choice experiment approach was easy to administer among a frail population so probably should not be ruled out on these grounds in considering its application to other sectors. The approach provides a useful way of obtaining data from a relatively small number of patients and there is no need for follow-up. For populations where follow-up is a problem, this is a positive characteristic of this technique.

The key issue is whether a complex intervention can be described in a meaningful way by its attributes. If the attributes are descriptions of aspects of care (as they were in the PDC study) then these need to be universally understood by both the respondents and the policy-makers to mean the same thing. The choice experiment uses complex analytical tools to derive values that are easily understood by policy-makers. There is no reason why these tools might not be applied to wider social research contexts. However, the extent to which the values derived from choice experiments reflect the actual preferences (and the preferences of the appropriate people) in any given context is not generally known. Given the complexity of the analysis, and the importance of understanding the meaning behind the values that are derived, there is still a danger that this approach might be misused, as all evaluation techniques might. The particular problems found in this study, especially of identifying a quantitative variable, deriving WTP estimates and the associated problems of determining that respondents have adhere to the axioms of rationality in order for these WTP values to be valid, may not be confined to the evaluation of PDC.



**SUMMARY**

This chapter has reviewed the main findings of the empirical work and concluded that the preference-based approach provided more insightful evidence than the EQ-5D approach and answered questions. This is an important step towards a full economic evaluation that can reflect the value of these services to those who use them and provide evidence that is relevant to decision-makers. Lessons can be drawn from this to other areas of health and social care that face similar evaluation problems, and these are addressed in the concluding chapter of the thesis.

The final chapter draws together the arguments that have developed through the thesis, and considering the issues that are common to many complex services, and those that are specific to palliative day care. The specific contribution of the thesis to debates in health economic evaluation is highlighted alongside consideration of the limitations of the empirical work. It considers the contribution of the thesis to the theoretical debates outlined in the early chapters, and suggests an agenda for future work in this area.

## Chapter 8

**Conclusion****Introduction**

The aim of this thesis was to explore methods of measuring the outcomes of complex health care interventions within the conceptual framework of health economic evaluation. The example of palliative day care has offered the opportunity to explore the usefulness of current economic evaluation methodologies and to consider approaches that conceptualise the outcomes or benefits of health care interventions from the perspective of values and preferences rather than health gain. This chapter summarises the arguments and discussion set out in each chapter, tracing the main themes set out in the introduction as they have developed through the thesis.

This concluding chapter also reflects on the shortcomings of this empirical investigation. Finally the original contribution of the thesis is described in light of the aims of the thesis and areas for future research are outlined.

**SECTION 1. OVERVIEW OF THE MAIN ARGUMENTS DEVELOPED IN THE THESIS****The nature of palliative day care as an example of a complex intervention**

Chapter 1 introduced palliative day care as an umbrella term for services delivered in specialist day centres for people with life-limiting illness. PDC is predominantly used by people who live at home and who have been identified as having specific needs that can be met by a specialist palliative care service. These needs may be for symptom control, psychological support, social interaction, rehabilitation or other less easily defined needs that have to do with finding meaning, contentment, hope and self-control at the end of life. Palliative day care is a service that has developed in the voluntary sector as a response to these complex needs, using a range of different specialist and general services and support provided in a day centre that are meant to address or alleviate some of the problems people face living with advanced illness and facing up to dying.

Complex services have been defined in this thesis as interventions that aim to address a range of patients' (and others') needs, which may include clearly definable physical, psychological and social support, but also goals which are more subtle and indirect. These services aim to meet the different and changing needs in different ways, depending on a person's circumstances and experience. Therefore the notion of complexity relates to the

way people use the service (the range, timing and intensity of support offered) and to the heterogeneity of needs since the same intervention with the same aim and for the same objective 'problem' may not have the same effects on people. These services are structured to be flexible, individually focussed and to meet the subjective or felt needs of individuals.

Complexity also arises from the way in which services are organised. The range of services offered in a PDC centre is not the same in different day centres, and the emphasis on different aspects of care differs in subtle ways. Some centres offer every patient routine access to clinical specialists, whereas others emphasise the rehabilitative, creative or social aspects of care. These differences are not based on research evidence of what is effective in PDC but on the institutional values of the organisations that offer these services. There is still little evidence of how these differences in care patterns affect people's experience of care, or whether PDC meet their perceived needs.

Finally, PDC may be one of a number of services that patients may be accessing in a particular phase of their illness. They will usually have some form of specialist home nursing, attend specialist outpatient clinics and access primary care services. PDC is a relatively weak intervention since patients usually attend only once a week and other services will contribute to their overall welfare.

All of these characteristics of PDC are also present in other forms of complex interventions although not all these characteristics may be present in every one. The arguments set out in this thesis also apply to other services and the experience of economic evaluation of PDC services should provide some insights for researchers in these other fields. The challenge for economic evaluation is to try and make some progress in conceptualising and measuring the outcomes of these complex services in ways that can be incorporated into economic analysis. This can provide insights that will inform decisions about whether these services are cost-effective compared with other ways of caring for people, and compared with other calls on health and social care resources.

### **Welfare economic theory, the decision rules of economic evaluation and outcome measurement**

To consider which method of economic evaluation would provide the most valuable insights for research into complex interventions, chapter 2 took a step back into the discipline of economics to review the purpose of evaluation and the nature of the questions that economic evaluation tries to answer. The discussion in this chapter reviewed the foundations of economic evaluation, based on the theory of welfare economics. This theory has not

always been thought of as the best way of conceptualising the costs and benefits of policies. In the past, there has been some dissonance between those who thought that economics should be a normative social science concerned with how society ought to conduct itself, and those who wanted economics to develop natural laws akin to natural sciences and engineering. This dissonance has produced an intellectual schism in economic thought that has remained influential to this day. Pareto, whose laws of optimality are the foundations of modern welfare economics, was strongly in favour of economics being a positive rather than a normative science. Individuals were seen as the best judges of their own welfare, and social welfare could be maximised by individuals trading between alternative forms of consumption to maximise their individual utility. Normative judgement about how utility ought to be maximised could not be made; individuals were seen to be the best judges of value as observed through their decisions about wealth and consumption. This is the basis on welfare economics and health economics to this day. CBA is the practical application of welfare economics principles, but these principles have not always provided practical tools in contexts where monetary valuations cannot be easily obtained. Economic evaluation of public services such as health care has had to develop other methods of valuing the outcomes of policies in the face of the difficulties in applying CBA. The adoption of health outcome as a proxy for value has moved health economic evaluation away from the fundamental principles of welfare economics.

Chapter 2 went on to show that health economic evaluation has grappled with the problem of defining a theory of behaviour that can predict and explain the way that individuals behave, and respond to choice under uncertainty. The approaches to valuing the outcomes of health care that are commonly used violate at least some of the welfarist principles. The approach of measuring health gain has taken two distinct paths. One approach measures outcomes and then attempts to estimate the utility or value that consumers place on changes between health states before and after interventions. The other approach rejects some of the core principles of welfare economics in the context of health and health care, especially its focus on utility-based ideas of welfare. This approach defines the outcome as a quantity of health gain which has value of itself, thereby avoiding the problem of valuing health gain and measuring utility.

The objective valuation of the outcomes of health intervention used in both these approaches has largely replaced any notion of the financial valuation of outcomes. This has meant that health economic evaluation has worked well where changes in health can be defined unambiguously and objectively measured. For the evaluation of complex services, the nature of the intervention and the outcomes would suggest that such unambiguous

notions of health gain would be difficult to identify. Furthermore, since it is not known what these services contribute to overall welfare, health gain may only be one aspect of welfare among other domains of quality of life that the service contributes to. If this is the case, then a health gain approach will underestimate the overall contribution to welfare.

Health gain approaches are also problematic because they consider health gain separately from the value that individual's place on the service, or what they would sacrifice in order to consume health care. Since complex services aim to meet individual's needs, and these cannot be easily described in a health gain framework, other methods more akin to CBA methods, such as willingness-to-pay, might be more appropriate in these settings.

Methods of direct elicitation of WTP values have been strongly criticised for not reflecting real decision-making if there were a real market for health goods and services, and for being open to misunderstanding and political manipulation. Therefore, indirect approaches that consider how individuals would trade off specific attributes of health or health care against a price or other form of monetary sacrifice have been developed and introduced into evaluations. One of these methods, choice experimentation, uses a regression modelling technique to quantify the relative value of specific attributes and has been put forward as a way of obtaining indirect monetary valuations for CBA.

### **What can be learnt about measuring outcomes of complex services in the palliative care literature and in other related areas?**

Chapter 3 demonstrated that there has been almost no research in palliative care that specifically considers the value of changes in health-related quality of life to individuals. Studies have found no differences in HR-QoL between palliative care and 'standard' care but have recorded increased "satisfaction" with place of care by patients who receive palliative care. The studies that have reported these findings have not been undertaken primarily as economic evaluations but as clinical effectiveness or quality of life evaluations. The connection between patient satisfaction and consumer preferences has not been explicitly made and satisfaction has been reported in a way that suggests that this is only a secondary outcome.

Evidence from qualitative studies has suggested that patients do gain some benefits from PDC but this is still ambiguously defined. These studies were small scale and probably undertaken by palliative day care professionals keen to show the benefits of their service. These studies do, however, support the argument that the patients' perspective on the value of services would be more revealing than an objective measure of changes in health state.

Very little evidence could be found in other similar health care contexts, and this was related specifically to interventions for frail elderly people, and to evaluations of day hospital. These demonstrate that other interventions that are multidisciplinary, multidimensional, and individually focused have similar problems using the statutory methods of economic evaluation.

The research that has been published on palliative day care research has been sparse. The only evaluation study was the North Thames PDC study that found that the disease-specific approach to assessing effectiveness did not show any difference between patients who attended and patients who did not. Yet the evidence from open-ended interviews suggested that some patients who attended PDC expressed strong preferences for this service. Most patients said they had either changed or had a better outlook as a result of attending PDC. Therefore it was reasoned that the disease-specific approach to measuring outcome might not capture the aspects of quality of life that is important to patients.

This chapter demonstrated that no economic studies had been undertaken in palliative care that addresses the issue of measuring complex outcomes. There had however, been recent work undertaken to measure social outcomes for older people and this could represent an important step forward. However this study did not address all the issues of concern raised here since its focus was on the development of a single measure of outcome. As PDC offers a range of services, a study that found that PDC was 'effective' overall still could not answer the question 'what is it that makes PDC effective?' Since there is heterogeneity of service inputs, an approach to evaluation that uses outcome instruments could not provide answers to this question, or help guide policy towards the most effective forms of care.

### **Empirical investigation of the outcomes of PDC**

Chapter 4 reviewed the methods that would be used in the investigation of outcomes of PDC for this study. This was based on the empirical evidence reported in chapter 3 and the theoretical arguments set out in chapter 2. These suggested that a health gain approach might not be appropriate in the context of PDC, and that a preference-based approach might be the way forward in developing useful research tools in this setting.

Second, it was argued that small changes in quality of life might be perceived to be important to people in their last months or life, and who were not expecting to regain full health. Their expectations of achieving any improvement might be lower than in other phases of their lives, and the opportunity to improve in small ways might be appreciated more than the objective change in quality of life. This meant that an approach to economic

evaluation that incorporated strength of preference might be more suited to the problem of assessing the value of PDC.

Global quality of life instruments with narrow dimensions have succeeded in condensing the complexity of quality of life experiences into single measures of outcome. But if the dimensions of outcome of PDC cannot be easily captured by these dimensions, then *a priori* reasoning would suggest that the single outcome instrument would not reflect a more complex outcome. A global rather than disease-specific health-related quality of life instrument had not been tried out in an evaluation of PDC. The results of the North Thames PDC study suggested that a health-related quality of life approach might not be appropriate in the context of PDC, but the empirical evidence had not tested this hypothesis. Therefore it was decided that this would be tried out in a second study of PDC. This was the first empirical evaluation of palliative care to try the EQ-5D instrument.

The Chichester study was a more robust study design than the North Thames PDC study. If the EQ-5D study still showed no difference in outcome between PDC attenders and non-attenders over time this would be stronger proof that either there was no benefit from these services, or that the health gain approach was not the way to approach the problem. However, this power of the study was calculated on the palliative care outcome scale (POS) and not on the EQ-5D instrument. This was a weakness of the study. It illustrates the problem of determining no difference in an economic study undertaken alongside a clinical effectiveness study where the study is powered to detect clinical changes. It would have been possible to calculate an appropriate sample size for a given change in EQ-5D scores on particular domains that it could be hypothesised would be most likely to change, say, pain/symptoms and anxiety/depression. The calculation would need to ensure that the resulting confidence intervals were narrow enough to distinguish between quadrants on the cost-effectiveness plane. The sample size would need to be such that it could distinguish the important clinical and economic changes over time and between palliative care groups, and this would be around 420 patients in this study, based on a sample size calculation undertaken *post-hoc* using the mean and standard deviation values at the second interview.

To have an idea of the necessary sample size before the study began, it would be necessary to have some indication from the patients and from the providers of care of what would be a clinically important change in health related quality of life, and to try and map that change onto an EQ-5D profile. It has been argued in this thesis that clinically important changes to the patients may not have been detected by the EQ-5D instrument (for example, if patients continue to experience some, though lesser, problems over time in a particular domain), but

one of the limitations of this study is that this was not fully explored with either the patients or with the clinicians providing their care.

At the same time, an approach to evaluation that could incorporate the values that individuals placed on specific attributes of PDC was also considered. One relatively new methodology that is emerging in health economics is choice experiments. This approach to evaluation asks individuals to value the individual attributes of service, rather than evaluate the service as a whole. It is based on Lancasterian demand theory and random utility theory and uses regression models to estimate the relative importance of particular pre-defined characteristics of care in individuals' utility functions.

This approach had a number of attractive qualities for the evaluation of PDC. First, the methodology obtained relative values of different attributes of a service from individuals, in a way that mirrors real decision-making. Second, choice modelling was proposed as a way of estimating indirect willingness-to-pay values that could be incorporated into CBA analysis. Third, many of the problems of obtaining longitudinal health gain data were avoided since the impact of a service on quality of life over time was not estimated. Fourth, the study could be designed to look relatively simple, and, if the data could be collected, could produce sophisticated results from a few simple questions.

Choice modelling is based on theoretical axioms that assume that individuals make rational choices in order to maximise utility, have stable choices over time, and that they are able to trade between different levels of preferences. These axioms have been explored and the preliminary findings are that individuals are not observed to be rational utility maximisers but adopt ways of sidestepping the cognitive demands of complex decision-making. Nevertheless despite this, the evidence suggests that this approach does produce values that reflect people's underlying preferences for different aspects of care.

The choice modelling approach was attractive as a research methodology for investigating the outcomes of PDC. The reasons would also apply to other areas of health and social care that are multidimensional and hard to evaluate using the usual methods of economic evaluation. First, the decompositional approach meant that it was possible to investigate the importance of different characteristics or attributes of PDC. This could address the question of which attributes were relatively more important to patients (as opposed to service providers) and contributed more to their overall welfare. Second, the choice experiment was seen as a step towards conceptualising how a cost-benefit approach to evaluating the



service might be undertaken. This goal could not be met within this first study since only qualitative attributes were used in the experiment.

The hypothesis was that the investigation of the outcomes of PDC using the global quality of life measure (such as EQ-5D) would not provide useful results so the choice experiment was part of a continuous process of investigation of how to construct and measure the outcomes of PDC. The choice experiment was undertaken not only because it was argued that health gain could not be measured in a meaningful way, but also because this approach did not adequately reflect the qualitative responses and preferences for different aspects of care that were obtained in the first PDC study. The choice model approach could build on the interview data and a previous study (by HRD) that assessed the structure, process and outcomes of PDC, and construct an assessment of PDC that could measure the strength of preference for particular aspects of the service.

### **The methods of empirical investigation**

Chapter 5 set out in detail the methods of investigation that were employed in this thesis. The choice experiment was designed so that it would have the highest chance of providing insights into preferences for PDC. A limitation was that it could not include willingness-to-pay as an attribute (described in terms of a daily charge for attending a day centre for example) as this was perceived as a potential threat to collecting any useful data since there was a danger that patients would not take part if there was a charge to consider, or that this attribute would dominate their preferences and could jeopardised the entire project. However, the decision not to include a financial attribute meant that the strength of preference for different attributes of PDC could not be scaled using a single absolute scale of value.

There are two important aspects to this problem that have been explored in this thesis. The first is that for a choice experiment to produce willingness-to-pay estimates, certain strict and less strict economic axioms of rationality are assumed. This is not what drives the current interest in whether choice experiments adhere to the economic axioms of rationality (which is primarily whether CE adheres to random utility theory which describes intended behaviour). But it is a necessary condition that CE data adhere to these axioms for WTP estimates to reflect the value of health care interventions. The empirical investigation of these axioms in health economic research has only recently begun and little has been published in this area. However the consensus to date is that willingness-to-pay estimates in the choice experiments do reflect some magnitude of value, if not the actual revealed preference value of those characteristics. Qualitative work is also underway that checks the

estimated monetary valuations with respondents to consider their agreement with the values derived from their responses. The findings of this work will reveal more about whether these values can be interpreted as reflecting a monetary sacrifice that can be incorporated into cost-benefit analysis.

The second reason is more practical. Since choice experiments are relatively new it was not clear whether the approach would work in a palliative care setting or how much of a protest would be made about the inclusion of a WTP attribute. The idea of valuing the benefits of health care in monetary terms is problematic as it is conditional upon ability to pay and assumes that people are willing and able to value the characteristics of health care in this way. Both these assumptions are particularly problematic in a palliative care setting. The pilot study and discussions with clinicians in the pre-pilot phase of the study produced very negative responses to the inclusion of a monetary attribute, such as a financial charge for day care. Palliative care clinicians interpreted this as the introduction to patients of charges by the back door. They did not want to see this attribute included in the study as it could raise unnecessary concerns and anxiety among patients who participated in the study.

There was also a danger that the inclusion of a financial attribute would be the most important attribute in individuals' decision-making and they would always choose the scenario that did not include a charge. This would mean that they were not prepared to trade and the study would not be able to estimate the relative values of other attributes. Since the value of the other attributes was the focus of the study, it was seen as more important to have some data on the relative value of the attributes without monetary valuation than to risk obtaining no data at all. Therefore this first choice experiment in PDC focussed on the attributes and on the research process using this new methodology rather than on willingness-to-pay. If a quantitative and meaningful variable to represent willingness-to-pay can be identified in a future study, this would represent another step forward in this research.

### **The findings of the empirical studies**

Chapter 6 reported the findings of the empirical studies. As expected, the EQ-5D study did not reveal any differences in health-related quality of life between those who attended PDC and a historical control group who did not. The qualitative evidence reported that patients' preferences for aspects of care were not directly related to the domains that are measurable using EQ-5D. From this empirical evidence, the conclusion is that the global health-related quality of life instrument for measuring health gain is not likely to be a sensitive enough approach for measuring changes in quality of life associated with attending PDC.

The choice experiment study has shown that economic evaluation can provide both some alternative ways of looking at the output of complex services, and a framework for asking different questions from those posed in clinical and quality of life research. Some of these concern the overall priority that should be given to different aspects of palliative day care. These relate to how benefits derived from the service can be valued, what should be included in the range of services offered, and how to ensure that patients' needs and preferences for care are being met.

One of the key concerns of providers is whether PDC is a specialist intervention or whether patients would be equally happy with a bath and hairdressing as with more specialist services. The findings suggest that patients have stronger preferences for specialist physical therapies than for accessing medical services every visit.

### **Insights from the findings**

Chapter 7 considered the results of the empirical studies for policy and in terms of the use of the methodology in similar settings. There is evidence that patients do use PDC as a specialist service rather than attending for social reasons only. From a resource use perspective this is important since the intensity and range of specialist services is sometimes dependent on the availability of specialist volunteers as well as paid staff. The evidence suggests that these specialist therapy services (such as aromatherapy, massage and reflexology) should be perceived as core rather than peripheral services, and that financial resources should be set aside for these aspects of care as well as for medical support. This is not straightforward since these services do not have a strong evidence base and funders may be keener to support those medical and clinical services that have proven efficacy. The argument for the provision of services that are not (yet) evidence-based is difficult to support in a climate where clinical evidence and transparency has taken centre stage in health policy and local decision-making.

However, the aims of the PDC service are to address the complex needs of individuals. These needs may be met in a number of ways within a set palliative care budget. This approach begins to identify in a systematic way what those needs are from the point of view of the user of these services. It can identify what aspects of care they prefer and what aspects of care they are happier to forego in order to obtain their preferred package of care (and maximise their welfare). This is an evidence-based approach to service provision if it is accepted that these people do have particular needs but that these may be different depending on individual circumstances that change over time. From the palliative care professional's point of view, there may be a need for some acceptance that what they have

always believed is good for patients may not be the aspects of care that are most preferred by them. The idea that the individual is the best judge of their own welfare would seem to be a good starting point in the context of complex services of this nature.

The argument that has been made in this thesis is that a preference-based, welfarist approach is appropriate in the context of valuing the benefits of complex services. It is not the approach that has been used most commonly in health economics because it is difficult to elicit preferences and often unnecessary when other simpler measures as proxies for welfare will suffice. But in the context of services that aim to meet a range of needs of individuals, a value-based approach would seem to be the most appropriate.

The danger of comparison across sectors is that the benefits gained from PDC interventions may not be measurable in objective ways, so that PDC will not score well in any league table of benefits, however defined. An important empirical question to test in the general public is whether health and quality of life at the end of life is seen as important in a 'good' society. This moves the debate back towards the normative constructs of economics – 'how society ought to be'. The contention is that palliative care is seen as good service by the general public (who give charitably to hospices and other voluntary organisations who care for people with advanced illness) even though its benefits cannot easily be defined or measured. The problems of evaluation reviewed in here have illustrated the point that all approaches to the measurement of human experience are fallible, and that the focus on the scientific measurement and quantification of health outputs can lead to policy conclusions that may not reflect society's wishes.

## **SECTION 2. METHODOLOGICAL CONTRIBUTION OF THE THESIS**

The thesis has situated the problem of evaluating complex services within more general current debates about how to evaluate health care. It has been argued that these debates have been driven by a tension in health economic evaluation between theory and practical research. On the one hand, it is important to base research on a theory that can determine whether a policy is good for society (and to specify clearly what this means in terms of maximising social welfare), and on the other to undertake research that can answer real world problems in ways that people can understand and that answers the important questions for policy. If health gain can be seen as a proxy for the value of benefit derived from health gain, then the relationship between the satisfaction of preferences and objectively measurable health is not problematic. In areas where outcomes are not easy to define, the relationship between health gain and people's preferences may not be so direct. The consequence is that services that provide benefits in ways that cannot be measured

using the quality of life instruments currently available may lose out in the competition for resources if they cannot demonstrate their value in other ways.

The review of the theory of economic evaluation and current approaches suggested that a preference-based approach to valuing health gain would provide more insightful evidence of the value of complex services. The choice experiment approach was considered and adopted in the empirical research undertaken, alongside a quality of life approach since this had not been tried in this setting.

The EQ-5D study did not show any difference in quality of life between patients who attended palliative day care and those who did not. The study also explored the specific attributes that are important to patients in a systematic way. If the only evidence that had been gathered had been from the open-ended interviews, then this would not have provided evidence of how much patients preferred some aspects of PDC to others. Specifically, the choice modelling data demonstrated that access to routine medical care was relatively less important (or less preferred) compared with other specialist services, and that bathing and hairdressing did not influence these patients' decisions to attend a palliative day care centre. The result is different from what could have been gained from interview data alone.

The choice experiment study demonstrated that these methods could be used in the context of palliative care and among patients who are mostly elderly, sometimes frail and possibly facing the end of their lives. When the study began, it was not clear that this method would be successful and, at the time, there was no evidence from other studies in similar contexts. The PDC study presented in this thesis provides evidence that these methods are possible to undertake with patients in a late phase of illness.

Further choice model experiments could consider outcomes in terms of how attributes are perceived by patients. For example, specialist therapies such as physiotherapy might allow patients to be more physically active and go out independently, or they may make people feel more positive about their illness. But the problem remains of how to describe these attributes in unambiguous ways, and how to express benefit using these criteria.

These are not easy questions to resolve. It requires more qualitative investigation of patients' perceptions of their ill health and how they see that they benefit from the care they receive. The argument that has been put forward in this thesis is that these questions about complexity require more complex strategies to answer them, and that simple methods using measurements of clinical or quality of life outcome designed for relatively straightforward

health care interventions do not provide the information necessary for decision-making. This is important to present as an argument when the current vogue is for standardised methodologies. For example, a recent article in the *Journal of the American Medical Association* has proposed that editorial teams and regulatory bodies use standardised instruments to assess the quality of economic evaluation research (Jefferson, Demyttenaere et al 2002). In the UK, the NICE has also proposed a standardised QALY approach for all health technology appraisals (Birch and Gafni 2002).

The study has also considered the extent to which it was possible to adopt a willingness-to-pay approach to measuring the outcomes of palliative day care, and whether a CBA approach would, in the future, be a possible way forward in PDC research. The direct WTP approach has been strongly criticised because of the problems of respondents equating WTP values with prices and not being sensitive to size of benefits in their valuation; that WTP valuations are made in costless and riskless decision-making contexts (without consequences of making a wrong valuation); and the potential problems of bias (by respondents acting to influence a decision, or by protest valuations). Choice modelling was proposed as a way of deriving indirect valuations for willingness-to-pay if a price was included among the attributes in the scenarios. However, the inclusion of a monetary attribute is still contentious in CE methodology, and its inclusion could jeopardise the whole research project. Therefore, this study only went part of the way towards a CBA approach to evaluating the outcomes of PDC.

### **Contribution to the palliative care literature**

Previous PDC studies have not attempted to consider the value of particular aspects of palliative care in a way that places these values at the centre of the evaluation methodology. This has not been because individuals' preferences were not seen as important (although they have not been considered important aspects of economic evaluation) but that the methods of incorporating values and preferences into evaluation have been relatively recent and may not have filtered into the palliative care research community. The added value of having a health economics perspective in evaluating PDC is that the concept of values and preferences can be placed at the core of the research paradigm.

This is the first study of its kind in palliative care that has been published (up until the end of 2002). The number of choice modelling experiments is increasing in the health sector. This study has used choice modelling experiments in a context where other methods of evaluating outcomes have not produced useful results. It is the first study to conceptualise and quantify the value of particular aspects of PDC and as such, can contribute to the

debate about how PDC should be organised. The method has proved to be popular in the day care centres where the data collection was undertaken, and it is hoped that these methods will be used in the context of other palliative care services that face similar evaluation issues.

### **Lessons to be drawn from the thesis for the evaluation of complex services**

The thesis has critiqued the use of health gain approaches for services that have a range of goals and where the main aim is to meet the needs of the individual. As the evaluation culture spreads to more services that cross the divide between health and social care, the health gain/cost-effectiveness/cost-utility paradigm may not provide insightful results in other complex care settings in the same way that they have not in palliative day care. Other less well-established methods such as choice modelling may need to be incorporated into the group of acceptable methods of economic evaluation by institutions such as NICE that currently favour cost-effectiveness and cost-utility methods. The debate will continue to be about when and how values should be incorporated into economic analyses, and how important they are. This study adds to this debate by questioning how PDC and similar services that aims to provide small but potentially valuable benefits can compete for resources if a strictly health gain approach is adopted.

The challenge of economic evaluation in palliative day care has highlighted problems that are common to other services that have similar characteristics to palliative day care, and that aim to be highly responsive to the individual needs of patients or clients who come into contact with the service. For services where the nature of the outcomes is hard to define and where they may be contested by different professional groups, an approach to evaluation that adopts a preference-based approach seems a useful way of obtaining evidence of the value of care and insights into contribution to welfare of specific aspects of the service. There may be a need in these circumstances to step away from the dominant paradigm in health economics evaluation and to challenge whether the current approaches that measure health gain in narrow dimensions are helpful in these contexts.

### **SECTION 3. SHORTCOMINGS OF THE STUDY DESIGN AND OF THE THESIS**

The thesis aimed to make progress in conceptualising and quantifying the outcomes of PDC in a way that could be incorporated into economic analysis. It was argued that subjective (value-based) rather than objective measures (of health gain) should be incorporated more explicitly into the evaluation of complex services. However, there are a number of drawbacks to this approach. First, the absolute value of the attributes of PDC could not be determined without the means of quantifying values or benefits in commensurate (monetary) units. It has

not been possible to define in a satisfactory way a unit of measurement to reflect what people would be prepared to give up in order to consume a specific attribute of PDC. This meant that the choice experiment could not be incorporated into a full cost-benefit analysis at this stage. Consequently, it was not possible to determine whether PDC was good value for money, compared with other calls on the same resources.

This does not mean that a CBA approach should be abandoned in the context of PDC, but that innovative methods of conceptualising sacrifice or trade-offs that could be translated into a monetary attribute could be explored. It has been suggested earlier that qualitative research could help in defining the nature of the sacrifice that patients may already make in order to attend PDC. As suggested earlier, it may be possible to determine what they might be prepared to give up in terms of other health and social care resources.

The fact that the other empirical studies provided no evidence of effectiveness does not mean these methods established evidence of no effectiveness. Nevertheless, it could be the case that PDC provides very marginal benefits (however defined) and that resources should be redirected to other services that can provide comparatively more benefits. The only supportive evidence that has been gathered from open-ended interviews suggests that PDC is valued at least by some people. However, not all patients made positive statements about PDC and there may be an important group of patients whose views were not incorporated in this study, namely those patients who attended PDC once and did not return, or chose not to attend at all.

The attributes that were defined in the choice modelling experiment were described in terms of the PDC inputs (and structure) rather than the putative effect that these have on patient outcomes. It has been argued in this thesis that describing the attributes of PDC in terms of outcomes would be difficult if not impossible, since the attributes might not be easy to describe in unambiguous language and simple descriptions of attributes may be interpreted differently. Standard gamble and other stated preference techniques require respondents to make trade-offs between health states or specific outcomes of health care. Therefore simple descriptions that reflected the essence of PDC would need to be made, and the argument presented here is that it would be difficult to get agreement on how these descriptions could be framed. Choice experiments have an advantage because they can allow health care interventions to be valued in terms of their characteristics or by attributes beyond health outcomes rather than in terms of changes in health state *per se*. These descriptions of PDC could be less problematic to define and agree upon as they could be framed in terms of physical characteristics of the service rather than their impact on the individuals who



experience them (which it is argued in this thesis, is different among individuals and subtle and complex by nature). This property of choice experiments may make this approach particularly useful in the evaluation of complex interventions.

However describing PDC in terms of outcomes has not been tested empirically. It would have been interesting to pilot a choice experiment that attempted to describe the attributes of day care in terms of specific outcomes or goals. This might provide information on what outcomes are important to patients overall, regardless of the palliative care setting (home, hospice, and hospital). However since the outcomes were complex, the decision to use this service-based approach in this first choice experiment study could be defended so long as the findings reflected the fact that the expressed preferences of respondents relate only to these pre-defined inputs.

The review of the theory of choice experiments has also demonstrated that the assumptions on which choice experiments are based are very strong, and that there is evidence that respondents do not consistently behave as rational welfare maximisers. Care needs to be taken not to over-interpret the results of these studies but to see them, like all research in these complex areas, as another form of compromise away from the theoretically correct model of evaluation. Preference-based approaches are closer to the core principles of welfare economic theory than health gain approaches, but are not a panacea for all evaluation problems.

### **The focus on patients' preferences in the choice experiment and not on carers, health professionals, or the wider public**

The study focused on the users of palliative day care and not a wider group of people who may have also have preferences that should be taken into account. The value to people other than the patients, say their carers, could be seen as an externality in the economic analysis of the benefits of the service. The reason that these externalities were not estimated was partly due to the experience of undertaking the North Thames PDC study prior to the study reported in this thesis. The North Thames study found that, while no difference in health-related quality of life was detected using palliative care specific quality of life instruments, interviews indicated that the users of the service reported benefits from attending and strong preferences for particular forms of care. The insight that users of the service had strong views of what they wanted was seen as an important starting point for the work presented in this thesis. Since the service is provided and funded to meet their needs, it is relevant to consider that their views should be the primary focus of the study to find out what services should be provided to meet their needs.

The focus of the choice experiment was to try and obtain more concrete evidence that PDC attenders have preferences for particular aspects of palliative day care, and that these could be identified and measured. The pilot test of the choice experiment questionnaire among public health students who were not familiar with the service suggested that they found it difficult to make the trade-offs required in the exercise since they did not know the value of these services to patients. Some of the students were highly reluctant to complete the task and gave this as their reason.

An argument can also be made that the wider public should be asked their preferences since these services are funded by voluntary contributions and taxation (through NHS contracts). However, if preferences for services that have not been directly experienced are not well formed (as for the public health students in the pilot study), there is a danger that this would become an exercise in guessing patients' preferences, rather than expressing those of a wider group of people, as also argued by Shiell and colleagues (Shiell et al 2000). The problem is also that without clear evidence of the effectiveness of PDC, the public does not have much more than prejudice and anecdote on which to make decisions between PDC and other forms of care. Furthermore the preferences of the general public might be that palliative care should meet the needs of those who use it. Therefore they are the ones who should be asked their views.

**If a CBA study was not undertaken, then is this method any better than the CEA/QALY approach?**

Both CUA and CEA can provide decision rules for the allocation of health care resources if one alternative dominates all others, or if there are commonly agreed upon threshold values for cost-effectiveness ratios. This thesis has argued that these approaches are not strictly welfarist since they both allow for interpersonal comparison of utility, and do not consider expressed values and preferences to be the only measures of utility, since health gain is measured separately from the value of health gain. However, they are equitable in that they assume that a unit of health is the same regardless of who receives it. A methodology that incorporates preferences is less equitable in the sense that it allows for the changes in health status to be more valuable for some than for others (those who would sacrifice or pay more for them). Whether this is an acceptable moral position for the allocation of resources is a strongly debated point, and goes to the heart of the earliest debates about normative and positive judgments in welfare economics.

Both approaches have their strengths and weaknesses. They are both valuable in making the process of decision-making more transparent. In contexts where the CEA and CUA

approaches do not provide any useful information to help make priority-setting decisions, other methods of evaluation are needed. The choice experiment has been able to provide insights into how PDC services are valued in ways that would not be possible using health gain approaches. It is not a complete picture, nor does it go far enough as to yield absolute values for the attributes of palliative day care. However the evidence that has been produced is a positive step along this road. Research methods that can capture some of this complexity contribute to knowledge where none (or very little) has previously existed.

#### **SECTION 4. AREAS OF FUTURE RESEARCH IN PALLIATIVE CARE**

This final chapter has signalled some particular areas where this research might be developed. These relate to the context of palliative care research (and similar areas), and to the design of the empirical research.

Economic studies in palliative care research are still dominated by the focus on evidence to demonstrate that palliative care can reduce costs. Studies have found no evidence of difference between conventional and palliative care, and this has been interpreted as evidence of no difference. The dominant economic evaluation methodology has been to undertake cost studies, assuming no difference in effectiveness. This approach has been able to avoid the problem of measuring complex outcomes.

The case has been put here that the value of particular outcomes or forms of care at the end of life may have more subjective value to the person experiencing the end of life than can be objectively measured using health-related quality of life instruments. Also, complex forms of care may also be less amenable to objective measurement due to the subtle nature of aims and outcomes of the service. If so, and there is something unique and different about the experience of life at the end of life compared with other periods of life, then providers of palliative day care and other complex services should welcome discussion about other methods of economic analysis that focus on patients' values and preferences, rather than an objective measure of outcome. This study has provided a novel way for the palliative care community to consider the value of their services, and it would be interesting to assess whether these methods could be used in other contexts where other evaluation methods have not been very successful. It is necessary to consider carefully whether patients who are more acutely ill and in deteriorating health would be able to participate in a choice modelling interview, but our research has shown that, contrary to expectations, this method of eliciting responses has not been onerous and has even been reported as enjoyable by some of the PDC respondents. The wider dissemination of our findings should hopefully precipitate some debate about the use of these methods in other palliative care studies.

The attributes chosen for this study may not have been the most important aspects of care to the patients who were interviewed, even though each attribute represented characteristics of care in the four participating centres. Further work to try and identify more sophisticated ways of expressing the outcomes of complex services is needed. This may require an in-depth qualitative study to investigate the experience of people who use these services as well as people who do not. Concepts such as “the social environment” need to be deconstructed into specific aspects of these broad characteristics in ways that make sense both to the respondents and to the providers of care so that they could organise their service more effectively.

It is important to make some more progress towards defining the concept of sacrifice in a choice experiment in a way that would be ethically and morally acceptable. Notions of sacrifice based on money or time are not seen as appropriate for people who are vulnerable in society or who may only have weeks or months to live. The unit of sacrifice would need to be a continuous variable against which the other attributes could be evaluated. Such a variable was not definable by the research team, even at the end of this study and after much thought. However, more in-depth analysis of this question might make more progress in this important area.

Finally, the method of validating the choice experiment (testing for rational consistent decision-making by respondents) was simple in this study. After the planning stages of the design of the experiment were undertaken, more sophisticated methods of validity testing have been proposed in the health economics literature. Research is still underway to test these methods of validation, but a future study should incorporate more robust methodological tests. Nevertheless, the preliminary reports of the findings of these validation studies suggest that the axioms of rationality underpinning choice modelling do not strongly hold. This is still an exploratory field in the health economics field. As long as the methodological issues are examined alongside the results of experiments, then choice modelling may have much to offer the evaluation of health sectors that are complex and where other methods have not provided practical information.

## **SECTION 5: THE CONTRIBUTION OF THIS RESEARCH TO ECONOMIC EVALUATION OF COMPLEX SERVICES**

A major part of the empirical analysis undertaken for this thesis has been to test the use of choice experiments in a sector that is inherently complex. It has been shown in this that empirical quantification of the strength of preferences of people who use these services is possible, despite the frailty of the respondents and despite the fact that palliative care is a

particularly sensitive area of health and social care due to its relationship with the dying process. This has been demonstrated even though it was not at all clear at the beginning of the process that it would be feasible and that we would be allowed access to interview these patients using a new research methodology in this field. Choice experiments can work and are acceptable to health care providers working with people who have a life-limiting illness and are (potentially) facing their death even though this is a sensitive area for outcome/valuation research and practice. A similar approach could be feasible in other sensitive areas of health and social care.

There are caveats to this argument since one of the criteria for attending palliative day care is that patients do not have acute mental health needs and are able to participate in and contribute to group and individual activities. The extent to which the experience of participating in a choice experiment exercise would be feasible among people with, say, dementia, has not been established. However, one of the contributions of this empirical work has been to provide evidence that choice experiment techniques can work in contexts that are inherently difficult, and has succeeded in gathering data despite some initial opposition from health care professionals. It argues that there is a wider range of contexts within which these methods might be used than perhaps the research and provider community has yet considered.

The wider application of choice experiment studies is now underway. A programme of research to consider how to incorporate 'sacrifice' into trade-offs in ways that could be compared across interventions would be an important contribution to applied empirical work. This study has not been able to follow all the necessary steps towards a full cost-benefit type study by not incorporating a quantifiable monetary measure (or proxy measure) for sacrifice. However, there may be ways of incorporating other notions of sacrifice as defined by the patients themselves into a CE study that would provide a way of comparing different kinds of services offered to this patient group. If a common unit of sacrifice could be identified, perhaps in terms of other services or resources people might be willing to give up in order to have a particular service, this would enhance the comparability of studies across other health and social care sectors. One way to do this, for patients with a range of inputs or services in their care package, would be to design broader choice experiments that would have the service of interest (say PDC, or community psychiatric nursing) as only one of a range of possible attributes that a respondent could choose from in a care package. If a respondent was prepared to sacrifice, say, home visits by a nurse for increased attendance at a palliative day care centre, then that would provide some notion of sacrifice that was relevant to the real situation the respondent faced, without incorporating a monetary

sacrifice. The opportunity cost of the resource that had been foregone could act as a proxy for value instead of market price or the maximum willingness-to-pay. This approach would not address the problems of valuing multidimensional interventions also addressed in this thesis (a palliative day care day or psychiatric nurse home visits would be taken to be a uniform entity), but this aspect of the research question would have to be traded off against the increased comparability of results of these studies. Certainly more innovative notions of individuals' economic sacrifice need to be developed to reflect the comparative value of complex interventions.

A broader perspective to explore 'society's' valuation of complex services is also timely. It is argued that the views of carers, of other health professionals (such as GPs, cancer specialists) and of the lay public should be incorporated in exercises to value these services. Approaches to incorporating these wider perspectives and preferences in economic evaluation of individual services are still not common. The consideration of whose values ought to count and how they should be incorporated into research is not always explicit in empirical research although it is more advanced in other sectors of economics.

The problem of evaluating complex services that do not fit easily into the mainstream health economic evaluation framework challenges the discipline in a number of ways. The analysis of the problem of economic evaluation of palliative day care highlights the problem of evaluating services that are inherently complex in terms of their inputs and outputs. It forces the researcher to consider what should be valued, how it should be valued, and by whom, and highlights some of the difficulties in evaluation that are not always addressed in empirical research. Therefore research in these areas can demonstrate some of the limitations in the mainstream health economic evaluation framework. This may be an important contribution of this thesis to the current body of health economic evaluation literature.

## Bibliography

- Abel, E.K. (1986). The hospice movement: institutionalizing innovation. *Int J Health Serv*, **16**, 71-85.
- Acito, F. (1977). An investigation of some data collection issues in conjoint measurement. In *Contemporary marketing thought*, Greenberg, B.A. & Bellenger, D.N. (eds) pp. 82-85. Educators Conference Proceedings: Chicago.
- Ackerman, F. (1979). An investigation of some data collection issues in conjoint measurement. In *Contemporary Marketing Thought*, Greenburg, B.A. & Nellenger, D.N. (eds). Educators Conference Proceedings: Chicago.
- Adamowicz, V. (2002). Design and analysis of choice experiments: lessons from environmental economics. *Discrete Choice/ Conjoint Analysis workshop*: Odense, Denmark.
- Aiken, L.H. (1986). Evaluation research and public policy: lessons from the National Hospice Study. *J Chronic Dis*, **39**, 1-4.
- Aksoy, S. (2000). Can the "quality of life" be used as a criterion in health care services? *Bull Med Ethics*, 19-22.
- Al, M.J., van Hout, B.A., Michel, B.C. & Rutten, F.F. (1998). Sample size calculation in economic evaluations. *Health Econ*, **7**, 327-35.
- Albinsson, L. & Strang, P. (2002). A palliative approach to existential issues and death in end-stage dementia care. *J Palliat Care*, **18**, 168-74.
- Altman, D.G. (1999). *Practical statistics for medical research*. Chapman and Hall: London.
- Amaya, M. (2002). Do individuals adopt non-compensatory decision-making? Preliminary evidence from health care. In *Discrete choice/ conjoint analysis workshop*: Odense, Denmark.
- Anderson, E. (1993). *Value in ethics and economics*. Harvard University Press: Cambridge, MA.
- Appleby, J. (1994). Contracting with the National Health Service: revised Guidelines for Voluntary Hospices. In *Occasional Paper*. National Council for Hospice and Specialist Palliative Palliative Care: London.
- Arnesen, T. & Nord, E. (1999). The value of DALY life: problems with ethics and validity of disability adjusted life years. *BMJ*, **319**, 1423-5.
- Arrow, K.J. (1963). Kenneth Arrow on welfare economics. *Am Econ Rev*, **53**, 941-973.

- Arrow, K., Solow, P., Portnoy, E., Leamer, E., R., R. & Schuman, H. (1993). *Report of the NOAA Panel on Contingent Valuation*. Federal Register 4601-4614.
- Avineri, S., and de-Shalit, A., (eds), 1992, *Communitarianism and Individualism*, Oxford: Clarendon Press
- Axelsson, B. & Christensen, S.B. (1998). Evaluation of a hospital-based palliative support service with particular regard to financial outcome measures. *Palliat Med*, **12**, 41-9.
- Backhouse, M.E. (2002) Use of randomised controlled trials for producing cost-effectiveness evidence: potential impact of design choices on sample size and study duration. *Pharmacoeconomics* **20**, 1061-77
- Badia, X., Herdman, M. & Kind, P. (1998). The influence of ill-health experience on the valuation of health. *Pharmacoeconomics*, **13**, 687-96.
- Badia, X., Roset, M., Montserrat, S., Herdman, M. & Segura, A. (1999). [The Spanish version of EuroQol: a description and its applications. European Quality of Life scale]. *Med Clin (Barc)*, **112** Suppl 1, 79-85.
- Bala, M.V., Wood, L.L., Zarkin, G.A., Norton, E.C., Gafni, A. & O'Brien, B. (1998). Valuing outcomes in health care: a comparison of willingness to pay and quality-adjusted life-years. *J Clin Epidemiol*, **51**, 667-76.
- Barclay, S., Todd, C., McCabe, J. & Hunt, T. (1999). Primary care group commissioning of services: the differing priorities of general practitioners and district nurses for palliative care services. *Br J Gen Pract*, **49**, 181-6.
- Baumgarten, M., Lebel, P., Laprise, H., Leclerc, C. & Quinn, C. (2002). Adult day care for the frail elderly: outcomes, satisfaction, and cost. *J Aging Health*, **14**, 237-59.
- Bayer, R., Callahan, D., Fletcher, J., Hodgson, T., Jennings, B., Monsees, D., Sieverts, S. & Veatch, R. (1983). The care of the terminally ill: morality and economics. *N Engl J Med*, **309**, 1490-4.
- Beggs, S., Cardell, S. & Hausman, J. (1981). Assessing the potential demand for electric cars. *Journal of Economics*, **16**.
- Bell, D.E. (1982). Regret in decision-making under uncertainty. *Operations Research*, **30**, 961-981.
- Bennett, K., G.M., T., M.H., B. & Guscott, R. (2000). Cost-utility analysis in depression: the McSad utility measure for depression health states. *Psychiatric Services*, **51**, 1171-1176.
- Billings, J. (1998). What is palliative care? *J Pall Med*, **1**, 73-81.



- Birch, S., Gafni, A. & O'Brien, B. (1999). Willingness to pay and the valuation of programmes for the prevention and control of influenza. *Pharmacoeconomics*, **16** Suppl 1, 55-61.
- Birch, S. & Gafni, A. (2002). On being NICE in the UK: guidelines for technology appraisal for the NHS in England and Wales. *Health Econ*, **11**, 185-91.
- Bjork, S. & Norinder, A. (1999). The weighting exercise for the Swedish version of the EuroQol. *Health Econ*, **8**, 117-26.
- Black, M. & Mooney, G. (2002). Equity in health care from a communitarian standpoint. *Health Care Anal*, **10**, 193-208.
- Black, R.D.C. (1990). Jevons, Marshall and the utilitarian tradition. *Scottish Journal of Political Economy*, **37**, 5-17.
- Blaug, M. (1985). *Economic theory in retrospect*. Cambridge University Press: Cambridge.
- Bleichrodt, H. (1997). Health utility indices and equity considerations. *J Health Econ*, **16**, 65-91.
- Bloom, B.S. & Kissick, P.D. (1980). Home and hospital cost of terminal illness. *Med Care*, **18**, 560-4.
- Blumenschein, K. & Johannesson, M. (1996). Incorporating quality of life changes into economic evaluations of health care: an overview. *Health Policy*, **36**, 155-66.
- Blumenschein, K. & Johannesson, M. (1999). Use of contingent valuation to place a monetary value on pharmacy services: an overview and review of the literature. *Clin Ther*, **21**, 1402-17; discussion 1401.
- Boardman, A.E., Greenberg, D.H., A.R., V. & Weimer, D. (2001). *Cost-benefit analysis. Concepts and practice*. Prentice-Hall: Englewood Cliffs, NJ.
- Bolmsjo, I., Hermeren, G. & Ingvar, C. (2002). Meeting existential needs in palliative care--who, when, and why? *J Palliat Care*, **18**, 185-91.
- Bonner, J. (1995). *Economic efficiency and social justice: the development of utilitarian ideas in economics from Betham to Edgeworth*. Edward Elgar: Aldershot, Hants.
- Bosanquet, N. (1997). New challenge for palliative care. *BMJ*, **314**, 1294.
- Bowling, A. (1995). What things are important in people's lives? A survey of the public's judgements to inform scales of health related quality of life. *Soc Sci Med*, **41**, 1447-62.
- Bowling, A. (1997). *Measuring health: a review of quality of life measurement scales*. Open University Press: Buckingham.

- Bowling, A. (2003). Adding quality to quantity: Older people's views on their quality of life and its enhancement. In *Growing Older: The ESRC Research programme on extending Quality of life*. [www.shef.ac.uk/uni/projects/gop/gop.3.html](http://www.shef.ac.uk/uni/projects/gop/gop.3.html): Economic and Social Research Council.
- Bradley, F., Wiles, R., Kinmonth, A.L., Mant, D., Gantley, M., , Group., T.S.C. & 13;318(7185):711-5., B.M. (1999). Development and evaluation of complex interventions in health services research: case study of the Southampton heart integrated care project (SHIP). *BMJ*, **318**, 711-5.
- Brady, M.J., Peterman, A.H., Fitchett, G., Mo, M. & Cella, D. (1999). A case for including spirituality in quality of life measurement in oncology. *Psychooncology*, **8**, 417-28.
- Brazier, J. & Dixon, S. (1995). The use of condition specific outcome measures in economic appraisal. *Health Econ*, **4**, 255-64.
- Brazier, J., Usherwood, T., Harper, R. & Thomas, K. (1998). Deriving a preference-based single index from the UK SF-36 Health Survey. *J Clin Epidemiol*, **51**, 1115-28.
- Brazier, J. & Deverill, M. (1999). A checklist for judging preference-based measures of health related quality of life: learning from psychometrics. *Health Econ*, **8**, 41-51.
- Brazier, J.E., Walters, S.J., Nicholl, J.P. & Kohler, B. (1996). Using the SF-36 and Euroqol on an elderly population. *Qual Life Res*, **5**, 195-204.
- Brazier, J., Deverill, M., Green, C., Harper, R. & Booth, A. (1999). A review of the use of health status measures in economic evaluation. *Health Technol Assess*, **3**, i-iv, 1-164.
- Brazier, J., Roberts, J. & Deverill, M. (2002). The estimation of a preference-based measure of health from the SF-36. *J Health Econ*, **21**, 271-92.
- Breitbart, W. (2002). Spirituality and meaning in supportive care: spirituality- and meaning-centered group psychotherapy interventions in advanced cancer. *Support Care Cancer*, **10**, 272-80.
- Briggs, A.H. & Gray, A.M. (1998). Power and sample size calculations for stochastic cost-effectiveness analysis. *Med Decis Making*, **18**, S81-92.
- Briggs, A.H., O'Brien, B.J. & Blackhouse, G. (2002). Thinking outside the box: recent advances in the analysis and presentation of uncertainty in cost-effectiveness studies. *Annu Rev Public Health*, **23**, 377-401.
- Briggs, A.H. & O'Brien, B.J. (2001). The death of cost-minimization analysis? *Health Econ*, **10**, 179-84.

- Bronowski, T. (1966). The logic of the mind. *The American Scholar*, **35**, 233-242.
- Broome, J. (1991). *Weighing goods: equality, uncertainty, time*. Blackwell: London.
- Broome, J. (1999). *Ethics out of economics*. Cambridge University Press: Cambridge.
- Brouwer, W.B. & Koopmanschap, M.A. (2000). On the economic foundations of CEA. Ladies and gentlemen, take your positions! *J Health Econ*, **19**, 439-59.
- Brown, C.V. & Jackson, P.M. (1978). *Public sector economics*. Blackwell: Oxford.
- Brown, J. & Sculpher, M. (1999). Benefit valuation in economic evaluation of cancer therapies. A systematic review of the published literature. *Pharmacoeconomics*, **16**, 17-31.
- Bruera, E. & Suarez-Almazor, M. (1998). Cost effectiveness in palliative care. *Palliat Med*, **12**, 315-6.
- Bryan, S., Buxton, M., Sheldon, R. & Grant, A. (1998). Magnetic resonance imaging for the investigation of knee injuries: an investigation of preferences. *Health Econ*, **7**, 595-603.
- Bryan, S., Gold, L., Sheldon, R. & Buxton, M. (2000). Preference measurement using conjoint methods: an empirical investigation of reliability. *Health Econ*, **9**, 385-95.
- Buckingham, K. (1993). A note on HYE (healthy years equivalent). *J Health Econ*, **12**, 301-9.
- Burrows, C. & Brown, K. (1992). Are any numbers better than no numbers? The sorry state of willingness-to-pay and some major methodological shortcomings. *Aust Health Rev*, **15**, 135-44.
- Burstrom, K., Johannesson, M. & Diderichsen, F. (2001). Swedish population health-related quality of life results using the EQ-5D. *Qual Life Res*, **10**, 621-35.
- Byford, S. & Sefton, T. (2002). First aid: lessons from health economics for economic evaluation in social welfare. LSE Health and Social Care: London.
- Cairns, J., van der Pol, M. & Lloyd, A. (2002). Decision making heuristics and the elicitation of preferences: being fast and frugal about the future. *Health Econ*, **11**, 655-8.
- Campen, J.T. (1986). *Benefit, cost and beyond*. Ballinger: New York.
- Carr-Hill, R.A. (1991). Allocating resources to health care: is the QALY (Quality Adjusted Life Year) a technical solution to a political problem? *Int J Health Serv*, **21**, 351-63.
- Cattin, P. & Wittink, D.R. (1982). Commercial use of conjoint analysis: a survey. *Journal of Marketing Research*, **46**, 44-53.

- Cave, M., Cunningham, D., Buxton, M., Hanney, S., Pollit, C., Scanlan, M. & Shurmer, M. (1993). *The valuation of changes in quality in the public services*. Vol. H.M. Treasury. HMSO: London.
- Cella, D.F. (1995). Measuring quality of life in palliative care. *Semin Oncol*, **22**, 73-81.
- Chan, S., Mackenzie, A. & Jacobs, P. (2000). Cost-effectiveness analysis of case management versus a routine community care organization for patients with chronic schizophrenia. *Arch Psychiatr Nurs*, **14**, 98-104.
- Chisholm, D., Healey, A. & Knapp, M. (1997). QALYs and mental health care. *Soc Psychiatry Psychiatr Epidemiol*, **32**, 68-75.
- Clark, D., Neale, B. & Heather, P. (1995). Contracting for palliative care. *Soc Sci Med*, **40**, 1193-202.
- Clarke, P.M. (2002). Testing the convergent validity of the contingent valuation and travel cost methods in valuing the benefits of health care. *Health Econ*, **11**, 117-27.
- Clinical-Standards-Advisory-Group. (1998). *Community health care for elderly people*. HMSO: London.
- Coast, J., Peters, T.J., Richards, S.H. & Gunnell, D.J. (1998). Use of the EuroQoL among elderly acute care patients. *Qual Life Res*, **7**, 1-10.
- Cohen, S.R., Mount, B.M., Bruera, E., Provost, M., Rowe, J. & Tong, K. (1997). Validity of the McGill Quality of Life Questionnaire in the palliative care setting: a multi-centre Canadian study demonstrating the importance of the existential domain. *Palliat Med*, **11**, 3-20.
- Cole, K., Cameron, J. & Edwards, C. (1983). *Why economists disagree*. Longman: London.
- Copp, G., Richardson, A., McDaid, P. & Marshall-Searson, D.A. (1998). A telephone survey of the provision of palliative day care services. *Palliat Med*, **12**, 161-70.
- Corr, C.A. & Corr, D.M. (1992). Adult hospice day care. *Death Stud*, **16**, 155-71.
- Craven, O. (2000). Palliative care provision and its impact on psychological morbidity in cancer patients. *Int J Palliat Nurs*, **6**, 501-7.
- Culyer, A.J. (1971). The nature of the commodity health care and its efficient allocation. *Oxford Economic Papers*, **24**, 189-211.
- Culyer, A.J. (1980) *The political economy of social policy*. Martin Robertson, Oxford.

- Culyer, A.J. (1990). Commodities, characteristics of commodities, characteristics of people, utilities and the quality of life. In *Quality of Life: perspectives and policies*, Baldwin, S., Godfrey, C. & Propper, C. (eds) pp. 9-27. Routledge: London.
- Culyer, A.J. & Wagstaff, A. (1993). QALYs (quality-adjusted life-years) versus HYEes (healthy years equivalents). *J Health Econ*, **12**, 311-23.
- Culyer, A.J. (1997). The rationing debate: Maximising the health of the whole community. The case for. *BMJ*, **314**, 667-9.
- Culyer, A.J. (2001). Economics and ethics in health care. *J Med Ethics*, **27**, 217-22.
- Cummings, J.E., Hughes, S.L., Weaver, F.M., Manheim, L.M., Conrad, K.J., Nash, K., Braun, B. & Adelman, J. (1990). Cost-effectiveness of Veterans Administration hospital-based home care. A randomized clinical trial. *Arch Intern Med*, **150**, 1274-80.
- Cummings, R.G., Brookshire, D.S. & Schulze, W.D. (1986). *Valuing environmental goods: a state of the arts assessment of the contingent method*. Rowman and Allanheld.: Totowa, N.J.
- Dasgupta, A.K. & Pearce, D.W. (1972). *Cost-benefit analysis: Theory and practice*. Macmillan Press: London.
- Dean, H.E. (1990). Political and ethical implications of using quality of life as an outcome measure. *Semin Oncol Nurs*, **6**, 303-8.
- Devlin, N., Hansen, P. & Herbison, P. (2000). Variations in self-reported health status: results from a New Zealand survey. *N Z Med J*, **113**, 517-20.
- Diener, A., O'Brien, B. & Gafni, A. (1998). Health care contingent valuation studies: a review and classification of the literature. *Health Econ*, **7**, 313-26.
- Dolan, P., Gudex, C., Kind, P. & Williams, A. (1996). Valuing health states: a comparison of methods. *J Health Econ*, **15**, 209-31.
- Dolan, P. (1999). Valuing health-related quality of life. Issues and controversies. *Pharmacoeconomics*, **15**, 119-27.
- Dolan, P. (2000). The measurement of health related quality of life. In *Handbook of health economics*, Culyer, A.J. & Newhouse, J.P. (eds), Vol. 1B. pp. Chapter 33. Elsevier Science: Oxford.
- Dolan, P. & Cookson, R. (2000). A qualitative study of the extent to which health gain matters when choosing between groups of patients. *Health Policy*, **51**, 19-30.
- Dolan, P. (2001). Utilitarianism and the measurement and aggregation of quality-adjusted life years. *Health Care Anal*, **9**, 65-76.

- Dolan, P. & Edlin, R. (2002). Is it really possible to build a bridge between cost-benefit analysis and cost-effectiveness analysis? *J Health Econ*, **21**, 827-43.
- Donaldson, C., Atkinson, A., Bond, J. & Wright, K. (1988). QALYS and long-term care for elderly people in the UK: scales for assessment of quality of life. *Age Ageing*, **17**, 379-87.
- Donaldson, C. (1990). Willingness to pay for publicly-provided goods. A possible measure of benefit? *J Health Econ*, **9**, 103-18.
- Donaldson, C. & Shackley, P. (1997). Does "process utility" exist? A case study of willingness to pay for laparoscopic cholecystectomy. *Soc Sci Med*, **44**, 699-707.
- Donaldson, C. (1999). Valuing the benefits of publicly-provided health care: does 'ability to pay' preclude the use of 'willingness to pay'? *Soc Sci Med*, **49**, 551-63.
- Donaldson, C. (2001). Eliciting patients' values by use of 'willingness to pay': letting the theory drive the method. *Health Expect*, **4**, 180-8.
- Donaldson, C., Currie, G. & Mitton, C. (2002). Cost effectiveness analysis in health care: contraindications. *BMJ*, **325**, 891-4.
- Douglas, H.R., Higginson, I.J., Myers, K. & Normand, C. (2000). Assessing structure, process and outcome in palliative day care: a pilot study for a multicentre trial. *Health Soc Care Community*, **8**, 336-344.
- Douglas, H.-R. & Normand, C.E. (2001). The economics of palliative day care. In *Palliative Day Care in Practice*, Hearn, J. & Myers, K. (eds). Oxford University Press: Oxford.
- Dowie, J. (2002). Decision validity should determine whether a generic or condition-specific HRQOL measure is used in health care decisions. *Health Econ*, **11**, 1-8.
- Doyle, D. (1992). Have we looked beyond the physical and psychosocial? *J Pain Symptom Manage*, **7**, 302-11.
- Dreze, J. (1962). [From the value of life to the economics and ethics of population: the path is purely methodological]. *Recherches Economiques de Louvain*, **58**, 147-166.
- Drummond, M.F., Mohide, E.A., Tew, M., Streiner, D.L., Pringle, D.M. & Gilbert, J.R. (1991). Economic evaluation of a support program for caregivers of demented elderly. *Int J Technol Assess Health Care*, **7**, 209-19.
- Drummond, M.F., O'Brien, B., Stoddart, G.L. & Torrance, G.W. (1997). *Methods for the economic evaluation of health care programmes*. Oxford Medical Publications. Oxford University Press: Oxford.

- Dunt, D.R. & Cantwell, A.M. (1989). the cost-effectiveness of the City Mission Hospice programme, Melbourne. *Palliat Med*, **3**, 125-134.
- Edwards, A., Daley, A. & Livingstone, H. (1997). Single-question assessment of quality of life. *Palliat Med*, **11**, 325-6.
- Edwards, R.T. (2001). Paradigms and research programmes: is it time to move from health care economics to health economics? *Health Econ*, **10**, 635-49.
- Efficace, F. & Marrone, R. (2002). Spiritual issues and quality of life assessment in cancer care. *Death Stud*, **26**, 743-56.
- Elixhauser, A., Luce, B.R., Taylor, W.R. & Reblando, J. (1993). Health care CBA/CEA: an update on the growth and composition of the literature. *Med Care*, **31**, JS1-11, JS18-149.
- Elixhauser, A., Halpern, M., Schmier, J. & Luce, B.R. (1998). Health care CBA and CEA from 1991 to 1996: an updated bibliography. *Med Care*, **36**, MS1-9, MS18-147.
- Elkan, R., Kendrick, D., Dewey, M., Hewitt, M., Robinson, J., Blair, M., Williams, D. & Brummell, K. (2001). Effectiveness of home based support for older people: systematic review and meta-analysis. *BMJ*, **323**, 719-25.
- Emanuel, E.J. (1996). Cost savings at the end of life. What do the data show? *JAMA*, **275**, 1907-14.
- Engquist, C.L. (1979). Can quality of life be evaluated? *Hospitals*, **16**, 97-100.
- Etzioni, A. (1993). *The spirit of community*. Crown Publishers: New York.
- EuroQol, G. (1990). EuroQol--a new facility for the measurement of health-related quality of life. The EuroQol Group. *Health Policy*, **16**, 199-208.
- Evans, R., 3rd, LeBailly, S., Gordon, K.K., Sawyer, A., Christoffel, K.K. & Pearce, B. (1999). Restructuring asthma care in a hospital setting to improve outcomes. *Chest*, **116**, 210S-216S.
- Eve, A. & Smith, A.M. (1994). Palliative care services in Britain and Ireland--update 1991. *Palliat Med*, **8**, 19-27.
- Evers, S.M., Van Wijk, A.S. & Ament, A.J. (1997). Economic evaluation of mental health care interventions. A review. *Health Econ*, **6**, 161-77.
- Farrar, S. & Ryan, M. (1999). Response-ordering effects: a methodological issue in conjoint analysis. *Health Econ*, **8**, 75-9.
- Farrar, S., Ryan, M., Ross, D. & Ludbrook, A. (2000). Using discrete choice modelling in priority setting: an application to clinical service developments. *Soc Sci Med*, **50**, 63-75.

- Faulkner, A., Higginson, I.J., Heulwen, E. & al, e. (1993). *Hospice day care: a qualitative study*. Help the Hospices: Sheffield.
- Feeny, D., Furlong, W., Barr, R.D., Torrance, G.W., Rosenbaum, P. & Weitzman, S. (1992). A comprehensive multiattribute system for classifying the health status of survivors of childhood cancer. *J Clin Oncol*, **10**, 923-8.
- Feeny, D., David, H., Torrance, G.W. & Furlong, W.J. (1996). Health Utilities Index. In *Quality of life and pharmacoeconomics*, Spilker, B. (ed). Lippincott-Raven Press: Philadelphia.
- Fisher, R.A. & McDaid, P. (1996). *Palliative day care*. Arnold: London.
- Forster, A., Young, J. & Langhorne, P. (1999). Systematic review of day hospital care for elderly people. The Day Hospital Group. *BMJ*, **318**, 837-41.
- Frank, R.H. (1989). Frames of reference and quality of life. *American Economic Review*, **79**, 80-89.
- Franks, P.J., Salisbury, C., Bosanquet, N., Wilkinson, E.K., Lorentzon, M., Kite, S., Naysmith, A. & Higginson, I.J. (2000). The level of need for palliative care: a systematic review of the literature. *Palliat Med*, **14**, 93-104.
- Furlong, W.J., Feeny, D.H., Torrance, G.W. & Barr, R.D. (2001a). The Health Utilities Index (HUI) system for assessing health-related quality of life in clinical studies. *Ann Med*, **33**, 375-84.
- Furlong, W.J., Feeny, D.H., Torrance, G.W. & Barr, R.D. (2001b). The health utilities Index (HUI) system for assessing health-related quality of life in clinical studies. *Annals of Medicine*.
- Gafni, A., Birch, S. & Mehrez, A. (1993). Economics, health and health economics: HYE's (healthy-years equivalent) versus QALYs (quality-adjusted live-year). *J Health Econ*, **12**, 325-39.
- Gafni, A. & Birch, S. (1995). Preferences for outcomes in economic evaluation: an economic approach to addressing economic problems. *Soc Sci Med*, **40**, 767-76.
- Gafni, A. (1997). Alternatives to the QALY measure for economic evaluations. *Support Care Cancer*, **5**, 105-11.
- Gafni, A. (1997). Willingness to pay in the context of an economic evaluation of healthcare programs: theory and practice. *Am J Manag Care*, **3 Suppl**, S21-32.
- Gafni, A. & Birch, S. (1997). QALYs and HYE's: spotting the difference. *Journal of Health Econ*, **16**, 601-608.



- Gafni, A. (1998). Willingness to pay. What's in a name? *Pharmacoeconomics*, **14**, 465-70.
- Gerard, K. (1988). An appraisal of the cost-effectiveness of alternative day care settings for frail elderly people. *Age Ageing*, **17**, 311-8.
- Gerard, K. (1992). Cost-utility in practice: a policy maker's guide to the state of the art. *Health Policy*, **21**, 249-79.
- Gerard, K., Smoker, I. & Seymour, J. (1999). Raising the quality of cost-utility analyses: lessons learnt and still to learn. *Health Policy*, **46**, 217-38.
- Gerard, K. (2002). Respondent consistency in discrete choice experiments: moving things forward using a case study in emergency and on demand health care systems. In *Discrete choice/ conjoint analysis workshop*: Odense, Denmark.
- Gibb, S., Donaldson, C. & Henshaw, R. (1998). Assessing strength of preference for abortion method using 'willingness to pay': a useful research technique for measuring values. *J Adv Nurs*, **27**, 30-6.
- Gigerenza, G. & D., G. (1999). Betting on one good reason: the "take the best" heuristic. In *Simple heuristics that make us smart*, Gigerenza, G. & Todd, P. (eds). Oxford University Press: New York.
- Glick, H.A., Polsky, D., Willke, R.J. & Schulman, K.A. (1999). A comparison of preference assessment instruments used in a clinical trial: responses to the visual analog scale from the EuroQol EQ-5D and the Health Utilities Index. *Med Decis Making*, **19**, 265-75.
- Goddard, M. (1989). The role of economics in the evaluation of hospice care. *Health Policy*, **13**, 19-34.
- Goddard, M.K. (1993). The importance of assessing the effectiveness of care: the case of hospices. *Journal of Social Policy*, **22**, 1-17.
- Gold, M.R., Siegel, J.E., Russell, E.M. & Weinstein, M.C. (1996). *Cost-effectiveness in health and medicine*. Oxford University Press: New York.
- Goodwin, D.M., Higginson, I.J., Edwards, A.G., Finlay, I.G., Cook, A.M., Hood, K., Douglas, H.R. & Normand, C.E. (2002a). An evaluation of systematic reviews of palliative care services. *J Palliat Care*, **18**, 77-83.
- Goodwin, D.M., Higginson, I.J., Myers, K., Douglas, H.R. & Normand, C.E. (2002b). What is palliative day care? A patient perspective of five UK services. *Support Care Cancer*, **10**, 556-62.

- Gosden, T., Bowler, I. & Sutton, M. (2000). How do general practitioners choose their practice? Preferences for practice and job characteristics. *J Health Serv Res Policy*, **5**, 208-13.
- Grande, G.E. & Todd, C.J. (2000a). Why are trials in palliative care so difficult? *Palliat Med*, **14**, 69-74.
- Grande, G.E., Todd, C.J., Barclay, S.I. & Farquhar, M.C. (2000b). A randomized controlled trial of a hospital at home service for the terminally ill. *Palliat Med*, **14**, 375-85.
- Gray, A.M., Marshall, M., Lockwood, A. & Morris, J. (1997). Problems in conducting economic evaluations alongside clinical trials. Lessons from a study of case management for people with mental disorders. *Br J Psychiatry*, **170**, 47-52.
- Green, P.E. & Krieger, A.M. (1991). Segmenting markets with conjoint analysis. *Journal of Marketing*, **55**, 20-31.
- Green, P.E. & Rao, V.R. (1971). Conjoint measurement for quantifying judgemental data. *Journal of marketing research*, **VIII**, 355-63.
- Green, P.E. & Wind, Y. (1975). New ways of measuring consumer judgement. *Harvard Business Review*, **53**, 107-117.
- Green, P.E. & Srinivasan, V. (1978). Conjoint analysis in consumer research: issues and outlook. *Journal of Consumer Research*, **5**, 103-123.
- Green, P.E. & Srinivasan, V. (1987). Conjoint analysis in consumer research: issues and outlook. *Journal of Consumer Research*, **5**, 338-376.
- Green, P.E., Tull, D.S. & Albaum, G. (1988). *Research for marketing decisions*. Prentice Hall: New Jersey.
- Green, P.E. & Srinivasan, V. (1990). Conjoint analysis in marketing: new developments with implications for research and practice. *Journal of Marketing*, **4**, 3-19.
- Greene, W. (2003). *Econometric analysis*. 5th Edition. Prentice-Hall: New York.
- Greer, D.S., Mor, V., Morris, J.N., Sherwood, S., Kidder, D. & Birnbaum, H. (1986). An alternative in terminal care: results of the National Hospice Study. *J Chronic Dis*, **39**, 9-26.
- Grieve, R., Porsdal, V., Hutton, J. & Wolfe, C. (2000). A comparison of the cost-effectiveness of stroke care provided in London and Copenhagen. *Int J Technol Assess Health Care*, **16**, 684-95.

- Gulliford, M.C. (1997). Design of cost-effective packages of care for non-insulin-dependent diabetes mellitus. Defining the information needs. *Int J Technol Assess Health Care*, **13**, 395-410.
- Haaijer, R. & Wedel, M. (2000). Conjoint choice experiments: general characteristics and alternative model specification. In *Conjoint measurement - methods and applications*, Gustafsson, A., Herrmann, A. & Huber, F. (eds). Springer: Berlin.
- Hammond, P.J. (1987). Social science: the theory of the impossible? In *Arrow and the foundations of the theory of economic policy*, Feiwel, G.R. (ed). New York University Press: New York.
- Hanneman, W.M. (1994). Valuing the environment through contingent valuation. *Journal of Economic Perspectives*, **8**, 19-43.
- Harris, J. (1987). QALYfying the value of life. *J Med Ethics*, **13**, 117-23.
- Harris, J. (1988). Life: quality, value and justice. *Health Policy*, **10**, 259-66.
- Harvard-Law-Review-Editorial. (1992). Ask a silly question...: Contingent valuation of natural resource damage. *Harvard Law Review*, **105**, 1981-2000.
- Hearn, J. & Higginson, I.J. (1999). Development and validation of a core outcome measure for palliative care: the palliative care outcome scale. Palliative Care Core Audit Project Advisory Group. *Qual Health Care*, **8**, 219-27.
- Hearn, J. & Myers, K.M. (2001). *Palliative day care in practice*. Oxford University Press: Oxford.
- Heckman, J.J. & Willis, R.J. (1976). Estimation of a stochastic model of reproduction: an econometric approach. In *Household reproduction and consumption*, Terleckyj, N. (ed). Bureau of Economic Research: New York.
- Herrmann, A., Schmidt-Gallas, F. & Huber, F. (2000). Adaptive conjoint analysis: understanding the methodology and assessing reliability and validity. In *Conjoint measurement: method and applications*, Gustafsson, A., Herrmann, A. & Huber, F. (eds). Springer: New York.
- Herth, K.A. (2001). Development and implementation of a Hope Intervention Program. *Oncol Nurs Forum*, **28**, 1009-16.
- Hicks, J. (1975). The scope and status of welfare economics. *Oxford Economics Papers*, **27**, 307-26.
- Higginson, I. (1993). Palliative care: a review of past changes and future trends. *J Public Health Med*, **15**, 3-8.

- Higginson, I. (1999). Palliative care services in the community: what do family doctors want? *J Palliat Care*, **15**, 21-5.
- Higginson, I.J., Goodwin, D., Normand, C.E. & Douglas, H.R. (2000 unpublished). Palliative day care: what effect does it have on outcomes? A multicentre prospective evaluation. *Palliative Medicine*.
- Higginson, I.J., Hearn, J., Myers, K. & Naysmith, A. (2000). Palliative day care: what do services do? Palliative Day Care Project Group. *Palliat Med*, **14**, 277-86.
- Higginson, I.J., Goodwin, D., Douglas, H.-R. & Normand, C.E. (2002). Evaluation of a new palliative day care service in Chichester. South Thames NHS Regional Office: London.
- Hinshaw, D.B., Carnahan, J.M. & Johnson, D.L. (2002). Depression, anxiety, and asthenia in advanced illness. *J Am Coll Surg*, **195**, 271-7; discussion 277-8.
- Hollingworth, W., Mackenzie, R., Todd, C.J. & Dixon, A.K. (1995). Measuring changes in quality of life following magnetic resonance imaging of the knee: SF-36, EuroQol or Rosser index? *Qual Life Res*, **4**, 325-34.
- Hopkinson, J.B. & Hallett, C.E. (2001). Patients' perceptions of hospice day care: a phenomenological study. *Int J Nurs Stud*, **38**, 117-25.
- Huber, F. (1987). Conjoint analysis: how we got here and where we are going. In *Sawtooth Software Research Paper*. Sequim: Washington.
- Huber, F. & Zwerina, K. (1996). The importance of utility balance in efficient choice designs. *Journal of Marketing Research*, **23**, 307-17.
- Huber, F., Herrmann, A. & Gustafsson, A. (2000). On the influence of the evaluation methods in conjoint design - some empirical results. In *Conjoint measurement - methods and applications*, Gustafsson, A., Herrmann, A. & Huber, F. (eds). Springer: Berlin, Germany.
- Hughes, S.L., Cummings, J., Weaver, F., Manheim, L., Braun, B. & Conrad, K. (1992). A randomized trial of the cost effectiveness of VA hospital-based home care for the terminally ill. *Health Serv Res*, **26**, 801-17.
- Hunt, E.K. (1979). Pure versus eclectic utilitarianism: the writing of Bastiat and Mill. In *History of economic thought: a critical perspective*. Wadsworth Publishing Company: Belmont, CA.
- Hurley, J. (1998). Welfarism, extra-welfarism and evaluative economic analysis in the health sector. In *Health, health care and health economics*, Barer, M.L., Getzen, T.E. & Stoddart, G.L. (eds). John Wiley: Chichester.

- Hutton, J. (1992). Cost-benefit analysis in health care expenditure decision-making. *Health Econ*, **1**, 213-6.
- Jackson, P. (1992). Welfare economics. In *What's new in economics?*, Maloney, J. (ed). Manchester University Press: New York.
- Jan, S. (1998). A holistic approach to the economic evaluation of health programs using institutionalist methodology. *Soc Sci Med*, **47**, 1565-72.
- Jan, S., Mooney, G., Ryan, M., Bruggemann, K. & Alexander, K. (2000). The use of conjoint analysis to elicit community preferences in public health research: a case study of hospital services in South Australia. *Aust N Z J Public Health*, **24**, 64-70.
- Jarvis, H., Burge, F.I. & Scott, C.A. (1996). Evaluating a palliative care program: methodology and limitations. *J Palliat Care*, **12**, 23-33.
- Jenkinson, C., Gray, A., Doll, H., Lawrence, K., Keoghane, S. & Layte, R. (1997). Evaluation of index and profile measures of health status in a randomized controlled trial. Comparison of the Medical Outcomes Study 36-Item Short Form Health Survey, EuroQol, and disease specific measures. *Med Care*, **35**, 1109-18.
- Johannesson, M. & Jonsson, B. (1991). Economic evaluation in health care: is there a role for cost-benefit analysis? *Health Policy*, **17**, 1-23.
- Johannesson, M. & Gerdtham, U. (1996). A note on the estimation of the equity-efficiency trade-off for QALYs. *J Health Econ*, **15**, 359-68.
- Johannesson, M. (1996). *Theory and methods of economics evaluation of health care*. Vol. 4. Kluwer Academic Press: Dordrecht, Sweden.
- Johannesson, M. (2001). Should we aggregate relative or absolute changes in QALYs? *Health Econ*, **10**, 573-7.
- Johnson, R. M. Banzhaf and W. Desvouges (2000) Willingness-to-pay for improved respiratory and cardiovascular health: a multiple format, stated-preference approach. *Health Economics*, **9**, 295-317
- Jones, A. (2001). *Applied econometrics for health economists - a practical guide*. Office of Health Economics: London.
- Jones-Lee, M.W. (1976). *The value of life: an economic analysis*. Martin Robertson: London.
- Jones-Lee, M.W., Hammerton, M. & Philips, P.R. (1985). The value of safety results of a national sample survey. *Economic Journal*, **95**, 49-72.

- Kaasa, S. & Loge, J.H. (2002). Quality-of-life assessment in palliative care. *Lancet Oncol*, **3**, 175-82.
- Kahneman, D., Slovic, P. & Tversky, A. (1982). *Judgement under uncertainty: heuristics and biases*. Cambridge University Press: Cambridge.
- Kahneman, D. & Knetsch, J. (1992). Valuing public goods: the purchase of moral satisfaction. *Journal of Environmental Management*, **22**, 57-70.
- Kane, R.L., Wales, J., Bernstein, L., Leibowitz, A. & Kaplan, S. (1984). A randomised controlled trial of hospice care. *Lancet*, **1**, 890-4.
- Kanninen, B. (2002). Optimal design of choice experiments for non-market valuation. In *Stated Preference: What do we know? Where do we go?* Environmental Policy and Economics: Washington, D.C., USA.
- Kaplan, R.M., Feeny, D. & Revicki, D.A. (1993). Methods for assessing relative importance in preference based outcome measures. *Qual Life Res*, **2**, 467-75.
- Kartman, B., Stalhammar, N.O. & Johannesson, M. (1996). Valuation of health changes with the contingent valuation method: a test of scope and question order effects. *Health Econ*, **5**, 531-41.
- Kawachi, I. (1989). QALYs and justice. *Health Policy*, **13**, 115-20.
- Kennedy, C.A. (2002). Revealed preference valuation compared to contingent valuation: radon-induced lung cancer prevention. *Health Econ*, **11**, 585-98.
- Kennett, C.E. (2000). Participation in a creative arts project can foster hope in a hospice day centre. *Palliat Med*, **14**, 419-25.
- Kilian, R. & Angermeyer, M.C. (1999). Quality of life in psychiatry as an ethical duty: from the clinical to the societal perspective. *Psychopathology*, **32**, 127-34.
- Kind, P., Dolan, P., Gudex, C. & Williams, A. (1998). Variations in population health status: results from a United Kingdom national questionnaire survey. *BMJ*, **316**, 736-41.
- Kirk, J. & Miller, M.L. (1985). *Reliability and validity in qualitative research*. Sage: Newbury Park, CA.
- Klose, T. (1999). The contingent valuation method in health care. *Health Policy*, **47**, 97-123.
- Krabbe, P.F., Essink-Bot, M.L. & Bonsel, G.J. (1997). The comparability and reliability of five health-state valuation methods. *Soc Sci Med*, **45**, 1641-52.
- Kroes, E.P. & Sheldon, R.J. (1988). Stated preference methods: an introduction. *Journal of transport economics*, **22**, 11-25.

- Lancaster, K. (1966). A new approach to consumer theory. *Journal of Political Economy*, **74**, 134-157.
- Lancaster, K. (1971). *Consumer demand: a new approach*. Columbia University Press: London.
- Lancet Editorial. (1986). Aggressive treatment of inoperable lung cancer: hoe much benefit for the patient? *Lancet*, **29**, 722.
- Landreth, H. & Colander, D.C. (1994). *The history of economic thought*. Houghton Mifflin Company: Boston.
- Lichtenstein, S. & Slovic, P. (1971). Reversal of preference between bids and choices in gambling decisions. *Journal of Experimental Psychology*, **89**, 46-55.
- Liljas, B. & Blumenschein, K. (2000). On hypothetical bias and calibration in cost-benefit studies. *Health Policy*, **52**, 53-70.
- Lindholm, L., Rosen, M. & Emmelin, M. (1998). How many lives is equity worth? A proposal for equity adjusted years of life saved. *J Epidemiol Community Health*, **52**, 808-11.
- Lohfeld, L.H., Tschopp, A.S., Trevor, A.W., Brazil, K. & Krueger, P. (2000). Assessing the need for and potential role of a day hospice: a qualitative study. *J Palliat Care*, **16**, 5-12.
- Longworth, L., Ratcliffe, J. & Boulton, M. (2001). Investigating women's preferences for intrapartum care: home versus hospital births. *Health Soc Care Community*, **9**, 404-13.
- Loomes, G. & Sugden, R. (1982). Regret theory: an alternative theory of rational choice under uncertainty. *Economic Journal*, **92**, 804-824.
- Loomes, G. & McKenzie, L. (1986). Disappointment and dynamic consistency in choice under uncertainty. *The Economic Journal*, **92**, 805-824.
- Loomes, G. & McKenzie, L. (1989). The use of QALYs in health care decision making. *Soc Sci Med*, **28**, 299-308.
- Loomes, G. (1995). The myth of the HYE (healthy year equivalent). *J Health Econ*, **14**, 1-7.
- Louviere, J. (1994). Conjoint analysis. Advanced methods of marketing research. In *Advanced methods of marketing research*, Bagozzi, R.P. (ed) pp. 223-259. Blackwell: Oxford.
- Louviere, J., Hensher, D. & Swait, J. (2000). Conjoint preference elicitation methods in the context of random utility theory preference elicitation methods. In *Conjoint*

- measurement - method and applications*, Gustafsson, A., Herrmann, A. & Huber, F. (eds). Springer: Berlin.
- Louviere, J., Hensher, R. & Swait, J. (2000). *Stated choice models: analysis and application*. Cambridge University Press: Cambridge.
- Louviere, J. (2001). Choice experiments: an overview of concepts and issues. In *The choice modelling approach to environmental valuation*, Bennett, J. & Blamey, R. (eds). Edward Elgar: Cheltenham (Northampton USA).
- Luce, R.D. & Tukey, J.W. (1964). Simultaneous conjoint measurement: a new type of fundamental measurement. *Journal of Mathematical Psychology*, **1**, 1-27.
- Ly, K.L., Chidgey, J., Addington-Hall, J. & Hotopf, M. (2002). Depression in palliative care: a systematic review. Part 2. Treatment. *Palliat Med*, **16**, 279-84.
- Mackay, A.F. (1980). *Arrow's theorem: the paradox of social choice: a case study in a philosophy of economics*. Yale University Press: London.
- Malhotra, N.K. (1982). Structural reliability and stability on nonmetric conjoint analysis. *Journal of Marketing Research*, **19**, 199-207.
- Manski, C. (1977). The structure of random utility models. *Theory and decisions*, **8**, 229-254.
- Marshall, M., Crowther, R., Almaraz-Serrano, A., Creed, F., Sledge, W., Kluiters, H., Roberts, C., Hill, E., Wiersma, D., Bond, G.R., Huxley, P. & Tyrer, P. (2001). Systematic reviews of the effectiveness of day care for people with severe mental disorders: (1) acute day hospital versus admission; (2) vocational rehabilitation; (3) day hospital versus outpatient care. *Health Technol Assess*, **5**, 1-75.
- Massaro, T. & McMillan, S. (2000). Instruments for assessing quality of life in palliative care settings. *Int J Palliat Nurs*, **6**, 429-33.
- Matthews, D.C., Gafni, A. & Birch, S. (1999). Preference based measurements in dentistry: a review of the literature and recommendations for research. *Community Dent Health*, **16**, 5-11.
- McCorkle, R., Benoliel, J.Q., Donaldson, G., Georgiadou, F., Moinpour, C. & Goodell, B. (1989). A randomized clinical trial of home nursing care for lung cancer patients. *Cancer*, **64**, 1375-82.
- McCrone, P., Beecham, J. & Knapp, M. (1994). Community psychiatric nurse teams: cost-effectiveness of intensive support versus generic care. *Br J Psychiatry*, **165**, 218-21.



- McCusker, J. & Stoddard, A.M. (1987). Effects of an expanding home care program for the terminally ill. *Med Care*, **25**, 373-85.
- McFadden, D. (1973). *Conditional logit analysis of qualitative choice behaviour*. University of California at Berkeley: California.
- McGuire, A., Henderson, J. & Mooney, G. (1988). *The economics of health care: an introductory text*. Routledge and Kegan Paul: London.
- McIntosh, E., Donaldson, C. & Ryan, M. (1999). Recent advances in the methods of cost-benefit analysis in healthcare. Matching the art to the science. *Pharmacoeconomics*, **15**, 357-67.
- McNiel, J., Stones, M. & Kozma, A. (1986). Subjective well-being in later life: issues concerning measurement and prediction. *Social indicators research*, **18**, 982-987.
- McWhinney, I.R., Bass, M.J. & Donner, A. (1994). Evaluation of a palliative care service: problems and pitfalls. *BMJ*, **309**, 1340-2.
- Mehrez, A. & Gafni, A. (1992). Preference based outcome measures for economic evaluation of drug interventions: quality adjusted life years (QALYs) versus healthy years equivalents (HYEs). *Pharmacoeconomics*, **1**, 338-45.
- Meister, H., Lausberg, I., Walger, M. & von Wedel, H. (2001). Using conjoint analysis to examine the importance of hearing aid attributes. *Ear Hear*, **22**, 142-50.
- Mishan, E.J. (1988). *Cost-benefit analysis*. Unwin Hyman Ltd: London.
- Mitchell, R.C. & Carson, R.T. (1989). *Using surveys to value public goods: the contingent valuation method*. Resources for the future: Washington D.C.
- Moayyedi, P., Wardman, M., Toner, J., Ryan, M. & Duffett, S. (2002). Establishing patient preferences for gastroenterology clinic reorganization using conjoint analysis. *Eur J Gastroenterol Hepatol*, **14**, 429-33.
- Mooney, G.H. (1977). *The valuation of human life*. Macmillan: London.
- Mooney, G. (1989). QALYs: are they enough? A health economist's perspective. *J Med Ethics*, **15**, 148-52.
- Mooney, G.H. & Olsen, J.A. (1991). QALYs: where next? In *Providing health care: the economics of alternative systems of finance and delivery*, McGuire, A., Fenn, P. & Mayhew, K. (eds). OUP: Oxford.
- Mooney, G. & Lange, M. (1993). Ante-natal screening: what constitutes 'benefit'? *Soc Sci Med*, **37**, 873-8.

- Mooney, G. (1994). What else do we want from our health services? *Soc Sci Med*, **39**, 151-4.
- Mooney, G. (1998). Beyond health outcomes. *Health Care Analysis*, **6**, 99-105.
- Mooney, G.H. (1997). Rationing and the objectives of health care. *Med J Aust*, **166**, 575.
- Mooney, G., Jan, S., Ryan, M., Bruggemann, K. & Alexander, K. (1999). What the community prefers, what it values, what health care it wants. A survey of South Australians. University of Sydney: Department of Public Health and Community Medicine.
- Mooney, G. (1998). "Communitarian claims" as an ethical basis for allocating health care resources. *Soc Sci Med*, **47**, 1171-80.
- Mooney, G. (2001). Communitarianism and health economics. In *The Social Economics of Health Care*, Davis, J.B. (ed). Routledge: London, New York.
- Morgan, A., Shackley, P., Pickin, M. & Brazier, J. (2000). Quantifying patient preferences for out-of-hours primary care. *J Health Serv Res Policy*, **5**, 214-8.
- Morrison, G.C. & Gyldmark, M. (1992). Appraising the use of contingent valuation. *Health Econ*, **1**, 233-43.
- Murray, C.J., Evans, D.B., Acharya, A. & Baltussen, R.M. (2000). Development of WHO guidelines on generalized cost-effectiveness analysis. *Health Econ*, **9**, 235-51.
- NAHAT. (1998). Promoting Partnership: Planning and managing community palliative care.: London.
- National Council for Hospice and Specialist Palliative Care Services. (1997). Working party on the future of palliative care. Occasional Paper Number 11, February: London.
- Neale, B. (1992). Palliative care in the community: the development of a rural hospice service in High Peak, North Derbyshire pp. 1-22. Trent Palliative Care Centre: Trent.
- Neale, B. & Clark, D. (1992). Informal care of people with cancer. *Journal of Cancer Care*, **1**, 193-198.
- Nease, R.F., Jr. (1996). Do violations of the axioms of expected utility theory threaten decision analysis? *Med Decis Making*, **16**, 399-403.
- Nelson, C.J., Rosenfeld, B., Breitbart, W. & Galletta, M. (2002). Spirituality, religion, and depression in the terminally ill. *Psychosomatics*, **43**, 213-20.
- Netten, A. & Curtis, L. (2002). Unit costs of health and social care. Personal and Social Services Research Unit, University of Kent: Canterbury.

- Netten, A., Smith, P., Healey, A. & Knapp, M. (2002). OPUS: A measure of social care outcome for older people. Discussion Paper 1690. Personal and Social Services Research Unit: University of Kent at Canterbury.
- Neumann, P.J., Stone, P.W., Chapman, R.H., Sandberg, E.A. & Bell, C.M. (2000). The quality of reporting in published cost-utility analyses, 1976-1997. *Ann Intern Med*, **132**, 964-72.
- Ng, Y.K. (1990). *Welfare economics: Introduction and development of Basic Concepts*. Macmillan Press: London.
- Nord, E. (1994). The QALY--a measure of social value rather than individual utility? *Health Econ*, **3**, 89-93.
- Nord, E., Richardson, J., Street, A., Kuhse, H. & Singer, P. (1995). Maximizing health benefits vs egalitarianism: an Australian survey of health issues. *Soc Sci Med*, **41**, 1429-37.
- Nord, E. (1995). The person-trade-off approach to valuing health care programs. *Med Decis Making*, **15**, 201-8.
- Nord, E. (1999). *Cost-value analysis in health care: making sense out of QALYs*. Cambridge University Press: Cambridge.
- Nord, E. (2001). The desirability of a condition versus the well being and worth of a person. *Health Econ*, **10**, 579-81.
- Normand, C. (1996). Economics and evaluation of palliative care. *Palliat Med*, **10**, 3-4.
- Norton, E.C. (2000). Long-term care. In *Handbook of health economics*, Culyer, A.J. & Newhouse, J.P. (eds), Vol. 1B. Elsevier Science: Oxford.
- O'Boyle, C.A. (1997). Measuring the quality of later life. *Philos Trans R Soc Lond B Biol Sci*, **352**, 1871-9.
- O'Brien, B. & Viramontes, J.L. (1994). Willingness to pay: a valid and reliable measure of health state preference? *Med Decis Making*, **14**, 289-97.
- O'Brien, B. & Gafni, A. (1996). When do the "dollars" make sense? Toward a conceptual framework for contingent valuation studies in health care. *Med Decis Making*, **16**, 288-99.
- Oliver, A. (2002). At the end of the beginning: eliciting cardinal values for health states. LSE Health and Social Care: London.
- Olsen, J.A. (1997). Aiding priority setting in health care: is there a role for the contingent valuation method? *Health Econ*, **6**, 603-12.

- Olshavsky, R.W. & Granbois, D.H. (1979). Consumer Decision-making - fact or fiction? *Journal of Consumer Research*, **6**, 93-100.
- Olson, S.L. (1989). Hospice day care standards development in Michigan. *The American Journal of Hospice Care*, **March/April**, 5-42.
- Onwujekwe, O., Chima, R., Shu, E., Nwagbo, D. & Okonkwo, P. (2001). Hypothetical and actual willingness to pay for insecticide-treated nets in five Nigerian communities. *Trop Med Int Health*, **6**, 545-53.
- Palmer, S., Byford, S. & Raftery, J. (1999). Economics notes: types of economic evaluation. *BMJ*, **318**, 1349.
- Patrick, D.L. & Deyo, R.A. (1989). Generic and disease-specific measures in assessing health status and quality of life. *Med Care*, **27**, S217-32.
- Payne, S.A., Langley-Evans, A. & Hillier, R. (1996). Perceptions of a 'good' death: a comparative study of the views of hospice staff and patients. *Palliat Med*, **10**, 307-12.
- Pearce, D.W. & Nash, C.A. (1981). *The social appraisal of projects*. Macmillan: London.
- Powers, D.A. & Xie, Y. (2000). *Statistical method for categorical data analysis*. Academic Press: San Diego, CA.
- Rabin, M. (1998). Psychology and economics. *Journal of Economic Literature*, **36**, 11-46.
- Racher, F.E., Kaufert, J.M. & Havens, B. (2000). Conjoint research interviews with frail elderly couples: methodological implications. *Journal of Family Nursing*, **6**, 367-379.
- Raftery, J.P., Addington-Hall, J.M., MacDonald, L.D., Anderson, H.R., Bland, J.M., Chamberlain, J. & Freeling, P. (1996). A randomized controlled trial of the cost-effectiveness of a district co-ordinating service for terminally ill cancer patients. *Palliat Med*, **10**, 151-61.
- Raiffa, H. (1968). *Decision analysis: Introductory lectures on choices under uncertainty*. Addison-Wesley: Reading, MA.
- Ratcliffe, J. & Buxton, M. (1999). Patients' preferences regarding the process and outcomes of life-saving technology. An application of conjoint analysis to liver transplantation. *Int J Technol Assess Health Care*, **15**, 340-51.
- Ratcliffe, J. & Longworth, L. (2002). Investigating the structural reliability of a discrete choice experiment within health technology assessment. *Int J Technol Assess Health Care*, **18**, 139-44.

- Rawls, J. *The theory of justice*. Harvard University Press. Cambridge, MA.
- Rinck, G.C., van den Bos, G.A., Kleijnen, J., de Haes, H.J., Schade, E. & Veenhof, C.H. (1997). Methodologic issues in effectiveness research on palliative cancer care: a systematic review. *J Clin Oncol*, **15**, 1697-707.
- Robbins, M.A. (1998). The economics of palliative care. In *Oxford Textbook of Palliative Medicine*, Doyle, D., Hanks, G. & MacDonald, N. (eds) pp. 55-65. Oxford University press: Oxford.
- Robinson. (1964). *Economic Philosophy*. Doubleday: Garden City, New York.
- Rosko, M.D., Walker, L.R., McKenna, W. & DeVita, M. (1983). Measuring consumer preferences for ambulatory medical care arrangements. *J Med Syst*, **7**, 545-54.
- Rosser, R. & Kind, P. (1978). A scale of valuations of states of illness: is there a social consensus? *Int J Epidemiol*, **7**, 347-58.
- Ryan, M. & Shackley, P. (1995). Assessing the benefits of health care: how far should we go? *Qual Health Care*, **4**, 207-13.
- Ryan, M. & Farrar, S. (1995). A pilot study using conjoint analysis to establish the views of users in the provision of orthodontic services in Grampian. In *Health Economists' Study Group Meeting*. University of Bristol.
- Ryan, M. (1996). Using consumer preferences in health care decision making: the application of conjoint analysis pp. 36. Office of Health Economics: London.
- Ryan, M. & Hughes, J. (1997). Using conjoint analysis to assess women's preferences for miscarriage management. *Health Econ*, **6**, 261-73.
- Ryan, M., McIntosh, E. & Shackley, P. (1998). Using conjoint analysis to elicit the views of health service users: an application to the patient health card. *Health Expect*, **1**, 117-129.
- Ryan, M. (1999). Using conjoint analysis to take account of patient preferences and go beyond health outcomes: an application to in vitro fertilisation. *Soc Sci Med*, **48**, 535-46.
- Ryan, M., McIntosh, E., Dean, T. & Old, P. (2000). Trade-offs between location and waiting times in the provision of health care: the case of elective surgery on the Isle of Wight. *J Public Health Med*, **22**, 202-10.
- Ryan, M., Bate, A., Eastmond, C.J. & Ludbrook, A. (2001). Use of discrete choice experiments to elicit preferences. *Qual Health Care*, **10 Suppl 1**, i55-60.

- Ryan, M. & Gerard, K. (2001). Using choice experiments to value health care programmes: current practice and future challenges. In *3rd International Health Economics Association Conference*: University of York, UK.
- Ryan, M., Scott, D.A., Reeves, C., Bate, A., van Teijlingen, E.R., Russell, E.M., Napper, M. & Robb, C.M. (2001). Eliciting public preferences for healthcare: a systematic review of techniques. *Health Technol Assess*, **5**, 1-186.
- Ryan, M. & Hughes, J. (2002). Listening to respondents : a think aloud study of discrete choice experiment responses. In *Health Economics Study Group Meeting*: University of East Anglia.
- Ryan, M. & San Miguel, F. (2002). Revisiting the axiom of completeness in health care. *Health Econ*.
- Sackett, D.L. & Torrance, G.W. (1978). The utility of different health states as perceived by the general public. *J Chronic Dis*, **31**, 697-704.
- Sacristan, J.A., Day, S.J., Navarro, O., Ramos, J. & Hernandez, J.M. (1995). Use of confidence intervals and sample size calculations in health economic studies. *Ann Pharmacother*, **29**, 719-25.
- Sagoff, M. (1994). Should preferences count? *Land Economics*, **2**, 127-145.
- Salter, J. (1994). Adam Smith on justice and distribution in commercial societies. *Scottish Journal of Political Economy*, **41**, 299-313.
- San Miguel, F. & Ryan, M. (2002a). Are preferences stable? the case of health care. *Journal of Economic Behaviour and Organization*, **48**, 1-14.
- San Miguel, F. & Ryan, M. (2002b). 'Rational' according to what? *Journal of Health Economics*, [published online August 2002].
- Sassi, F., Archard, L. & Le Grand, J. (2001). Equity and the economic evaluation of healthcare. *Health Technol Assess*, **5**, 1-138.
- Schelling, T.C. (1968). The life you save may not be your own. In *Public expenditure analysis*, Chase, S.B. (ed). Brookings Institute: Washington.
- Schwappach, D.L. (2002). Resource allocation, social values and the QALY: a review of the debate and empirical evidence. *Health Expect*, **5**, 210-22.
- Scitovsky, T. (1973). A note of welfare propositions in economics. In *Readings in welfare economics*, Farrell, M.J. (ed). St Martin's Press: London.
- Scitovsky, T. (1993). The meaning, nature and source of value in economics. In *The origins of value*, Hechter, M., Nadel, L. & Michod, R.R. (eds). Aldine de Gruyter: New York.

- Scott, A. (2001). Eliciting GPs' preferences for pecuniary and non-pecuniary job characteristics. *J Health Econ*, **20**, 329-47.
- Scottish Partnership Agency. (1995). *Day hospice: report of a meeting of the health service group*. Scottish Partnership Agency for Palliative Care and Cancer Care: Edinburgh.
- Sculpher, M., Gafni, A. & Watt, I. (2002). Shared treatment decision making in a collectively funded health care system: possible conflicts and some potential solutions. *Soc Sci Med*, **54**, 1369-77.
- Seale, C.F. (1989). What happens in hospices: a review of research evidence. *Soc Sci Med*, **28**, 551-9.
- Seale, C.F. (1991). A comparison of hospice and conventional care. *Soc Sci Med*, **32**, 147-52.
- Segal, M.N. (1982). Reliability of conjoint analysis: contrasting data collection procedures. *Journal of Marketing Research*, **13**, 211-224.
- Sen, A. (1970a). *Collective choice and social welfare*. Oliver and B.: London.
- Sen, A. (1970b). The impossibility of the Paretian liberal. *Journal of Political Economy*, **78**, 152-157.
- Sen, A. (1982). *Choice, welfare and measurement: resources, values and development*. Blackwell: Oxford.
- Sen, A. (1985). *Commodities and capabilities*. Oxford University Press: Dehli, India.
- Sen, A. (1987). *On ethics and economics*. Blackwell: Oxford.
- Shackley, P. & Ryan, M. (1995). Involving consumers in health care decision making. *Health Care Anal*, **3**, 196-204.
- Shackley, P., Slack, R. & Michaels, J. (2001). Vascular patients' preferences for local treatment: an application of conjoint analysis. *J Health Serv Res Policy*, **6**, 151-7.
- Shiell, A. (1997). Health outcomes are about choices and values: an economic perspective on the health outcomes movement. *Health Policy*, **39**, 5-15.
- Shiell, A., Hawe, P. & Seymour, J. (1997). Values and preferences are not necessarily the same. *Health Econ*, **6**, 515-8.
- Shiell, A., Seymour, J., Hawe, P. & Cameron, S. (2000). Are preferences over health states complete? *Health Econ*, **9**, 47-55.
- Shugan, S.M. (1980). The cost of thinking. *Journal of Consumer Research*, **7**, 99-111.
- Sintonen, H. (2001). The 15D instrument of health-related quality of life: properties and applications. *Ann Med*, **33**, 328-36.

- Skjoldborg, U.S. & Gyrd-Hansen, D. (2002). Conjoint analysis - the cost variable: an achilles' heel? *Health Econ.*
- Slovic, P. (1995). The construction of preferences. *American Psychologist*, **50**, 364-371.
- Small, K.A. & Rosen, H.S. (1981). Applied welfare economics with discrete choice models. *Econometrica*, **49**, 105-130.
- Smith, K.W., Avis, N.E. & S.F., A. (1999). Distinguishing between quality of life and health status in quality of life research: a meta-analysis. *Quality of Life Research*, **8**, 447-459.
- Smith, R.D. (1996). Is Regret Theory an alternative basis for estimating the value of healthcare interventions? *Health Policy*, **37**, 105-15.
- Smith, R.D. (2001). The relative sensitivity of willingness-to-pay and time-trade-off to changes in health status: an empirical investigation. *Health Econ*, **10**, 487-97.
- Smylie, J.A., Calvert, B.L. & Gerber, G.J. (1991). Program evaluation as a measure of the quality of a new approach to community psychiatric care. *Qual Assur Health Care*, **3**, 247-55.
- Spencer, D.J. & Daniels, L.E. (1998). Day hospice care--a review of the literature. *Palliat Med*, **12**, 219-29.
- Sugden, R. & Williams, A. (1978). *The principles and practice of cost-benefit analysis*. Oxford University Press: Oxford.
- Sullivan, S.D., Weiss, K.B., Lynn, H., Mitchell, H., Kattan, M., Gergen, P.J. & Evans, R. (2002). The cost-effectiveness of an inner-city asthma intervention for children. *J Allergy Clin Immunol*, **110**, 576-81.
- Swallow, S., Opaluch, J. & Weaver, T. (2001). Strength-of-preference indicators and an ordered-response model for ordinarily dichotomous discrete choice data. *J Environ Econ & Management*, **41**, 70-93.
- Taylor, W.J., Lord, S., McPherson, K.M. & McNaughton, H.K. (2001). EuroQol EQ-5D may not adequately describe the health of people with disabilities. *Disabil Rehabil*, **23**, 281-5.
- Tebbit, P. (1999). Palliative Care 2000. Commissioning through partnership. *National Council for Hospice and Specialist Palliative Care*: London.
- Thompson, B. (1990). Hospice day care. *The American Journal of Hospice Care*, **Jan/Feb**, 28-30.



- Tierney, A.J., Sladden, S., Anderson, J., King, M., Lapsley, I. & Llewellyn, S. (1994). Measuring the costs and quality of palliative care: a discussion paper. *Palliat Med*, **8**, 273-81.
- Tookman, A.J. & Scharpen von Heussen, K.S. (2001). The role of the doctor in day care. In *Palliative day care in practice*, Hearn, J. & Myers, K.M. (eds). Oxford University Press: Oxford.
- Torgerson, D.J. & Byford, S. (2002). Economic modelling before clinical trials. *BMJ*, **325**, 98.
- Torrance, G.W., Thomas, W.H. & Sackett, D.L. (1972). A utility maximization model for evaluation of health care programs. *Health Serv Res*, **7**, 118-33.
- Torrance, G.W. (1986). Measurement of health state utilities for economic appraisal. *J Health Econ*, **5**, 1-30.
- Torrance, G.W. & Feeny, D. (1989). Utilities and quality-adjusted life years. *Int J Technol Assess Health Care*, **5**, 559-75.
- Torrens, P.R. (1984). Studies of hospice economics. *Med Care*, **22**, 689-90.
- Torrens, P.R. (1985). Hospice care: what have we learned? *Annu Rev Public Health*, **6**, 65-83.
- Townend, M. (2000). An application of conjoint analysis to the process of psychiatric day hospital care. *Journal of Psych and Ment Hlth Nurs*, **7**, 371-372.
- Train, K. (1986). *Qualitative choice analysis. Theory, econometrics and an application to automobile demand*. Academic Press: Cambridge, MA.
- Tramarin, A., Milocchi, F., Tolley, K., Vaglia, A., Marcolini, F., Manfrin, V. & de Lalla, F. (1992). An economic evaluation of home-care assistance for AIDS patients: a pilot study in a town in northern Italy. *Aids*, **6**, 1377-83.
- Treadwell, J.R. & Lenert, L.A. (1999). Health values and prospect theory. *Med Decis Making*, **19**, 344-52.
- Tsuchiya, A. (2000). QALYs and ageism: philosophical theories and age weighting. *Health Econ*, **9**, 57-68.
- Tsuchiya, A., Ikeda, S., Ikegami, N., Nishimura, S., Sakai, I., Fukuda, T., Hamashima, C., Hisashige, A. & Tamura, M. (2002). Estimating an EQ-5D population value set: the case of Japan. *Health Econ*, **11**, 341-53.
- Tversky, A. & Kahneman, D. (1986). Rational choices for the framing of decisions. *Journal of Business*, **59**, 250-278.

- van den Hout, W.B., Van Der Linden, Y.M., Steenland, E., Wiggenraad, R.G., Kievit, J., De Haes, H. & Leer, J.W. (2003). Single- versus multiple-fraction radiotherapy in patients with painful bone metastases: cost-utility analysis based on a randomized trial. *J Natl Cancer Inst*, **95**, 222-9.
- van der Pol, M. & Ryan, M. (1995). Methodological issues in carrying out a conjoint analysis study: an application to consumer preferences for fruit and vegetables. *HERU Discussion Paper*, Department of Public Health Economics, University of Aberdeen.
- van der Pol, M. & Cairns, J. (1998). Establishing patient preferences for blood transfusion support: an application of conjoint analysis. *J Health Serv Res Policy*, **3**, 70-6.
- Ventafridda, V., De Conno, F., Vigano, A., Ripamonti, C., Gallucci, M. & Gamba, A. (1989). Comparison of home and hospital care of advanced cancer patients. *Tumori*, **75**, 619-25.
- Verhoef, L.C., de Haan, A.F. & van Daal, W.A. (1994). Risk attitude in gambles with years of life: empirical support for prospect theory. *Med Decis Making*, **14**, 194-200.
- Victor, C.R. & Higginson, I. (1994). Effectiveness of care for older people: a review. *Qual Health Care*, **3**, 210-6.
- Vinciguerra, V., Degnan, T.J., Budman, D.R., Brody, R.S., Moore, T., Sciortino, A. & O'Connell, M. (1986). Comparative cost analysis of home and hospital treatment. *Prog Clin Biol Res*, **216**, 155-64.
- von Neumann, J. & Morgenstern, O. (1944). *Theory of games and economic behaviour*. Princeton University Press: Princeton, NJ.
- Wagstaff, A. (1991). QALYs and the equity-efficiency trade-off. *J Health Econ*, **10**, 21-41.
- Wallston, K.A., Burger, C., Smith, R.A. & Baugher, R.J. (1988). Comparing the quality of death for hospice and non-hospice cancer patients. *Med Care*, **26**, 177-82.
- Ward, A. (1984). *Home Care Services for the terminally ill*. Nuffield Foundation: London.
- Warr, D. & Feld, R. (1996). The role of supportive care in quality of life. *Curr Opin Oncol*, **8**, 279-83.
- Webber, J. (1996). Demonstrating the value of palliative care. *Int J Pall Nurs*, **2**, 121.
- Weinstein, M.C., Siegel, J.E., Gold, M.R., Kamlet, M.S. & Russell, L.B. (1996). Recommendations of the panel on cost-effectiveness in health and medicine. *JAMA*, **276**, 1253-8.

- WHO. (1990). Cancer pain relief and palliative care. World Health Organisation, Geneva.
- WHOQOL\_Group. (1993). Study protocol for the World Health Organisation project to develop a quality of life assessment instrument. *Quality of Life Research*, 2, 153-159.
- Whynes, D. (1997). Costs of palliative care. In *New Themes in Palliative Care*, Clark, D., Hockley, J. & Ahmedzai, S. (eds). Oxford University Press: Oxford.
- Wigton, R.S., Hoellerich, V.L. & Patil, K.D. (1986). How physicians use clinical information in diagnosing pulmonary embolism: an application of conjoint analysis. *Med Decis Making*, 6, 2-11.
- Willan, A.R. & O'Brien, B.J. (1999). Sample size and power issues in estimating incremental cost-effectiveness ratios from clinical trials data. *Health Econ*, 8, 203-11.
- Willan, A.R. (2001). Analysis, sample size, and power for estimating incremental net health benefit from clinical trial data. *Control Clin Trials*, 22, 228-37.
- Willan, A.R., O'Brien, B.J. & Leyva, R.A. (2001). Cost-effectiveness analysis when the WTA is greater than the WTP. *Stat Med*, 20, 3251-9.
- Williams, A. (1988). Ethics and efficiency in the provision of health care. In *Philosophy and Medical Welfare*, Bell, M. & Mendux, S. (eds). Cambridge University Press: Cambridge.
- Williams, A. (1992). Priorities not needs! In *Meeting needs in an affluent society*, A., C., Robertson, E. & Trolley, K. (eds). Avebury: London.
- Williams, A. (1996). QALYS and ethics: a health economist's perspective. *Soc Sci Med*, 43, 1795-804.
- Williams, A. (1997). Intergenerational equity: an exploration of the 'fair innings' argument. *Health Econ*, 6, 117-32.
- Williams, A. (2000). Interpersonal comparisons of welfare. In *Health Economics Study Group*. Newcastle, UK.
- Williams, A. (2001). The 'fair innings argument' deserves a fairer hearing! Comments by Alan Williams on Nord and Johannesson. *Health Econ*, 10, 583-5.
- Wimo, A., Wallin, J.O., Lundgren, K., Ronnback, E., Asplund, K., Mattsson, B. & Krakau, I. (1990). Impact of day care on dementia patients--costs, well-being and relatives' views. *Fam Pract*, 7, 279-87.
- Winch, D. (1971). *Analytical welfare economics*. Penguin: London.

- Wordsworth, S., Ryan, M. & Waugh, N. (2001). Costs and benefits of cervical screening IV: valuation by women of the cervical screening programme. *Cytopathology*, **12**, 367-76.
- Zarnke, K.B., Levine, M.A. & O'Brien, B.J. (1997). Cost-benefit analyses in the health-care literature: don't judge a study by its label. *J Clin Epidemiol*, **50**, 813-22.
- Zhan, L. (1992). Quality of life: conceptual and measurement instruments. *J Adv Nurs*, **17**, 795-800.
- Zimmer, J.G., Groth-Juncker, A. & McCusker, J. (1984). Effects of a physician-led home care team on terminal care. *J Am Geriatr Soc*, **32**, 288-92.
- Zimmer, J.G., Groth-Juncker, A. & McCusker, J. (1985). A randomized controlled study of a home health care team. *Am J Public Health*, **75**, 134-41.
- Zwerina, K., Huber, F. & Kuhfield, W. (1996). *A general method for constructing efficient choice designs*. Fuqua School of Business, NC: Duke University.

**Appendix A Summary of the research evidence from a systematic review of economic studies in palliative care**

No.	Lead author, date and study methodology	Intervention	Viewpoint	Resource use data for	Measure of outcome	Results reported by authors
1	Vinciguerra (1986) Prospective case control	Home care vs hospital	Single institution	Hospital, home care	Pain, symptom control, satisfaction and ADL <sup>1</sup> .	Significantly higher scores for home care after two weeks for ADL and satisfaction. Home care produces 'significant savings'. Costs not reported alongside outcomes.
2	Cummings (1990) RCT	Home care vs conventional care	Multiple sectors	Hospital, private sector, ER, home care, private care (Patient records & diaries)	Self care, cognition, mental health, functioning, ADL, and satisfaction.	Significantly higher satisfaction with care, identical mean scores in other outcomes. Decrease in hospital and home care costs not associated with adverse outcomes.
3	Kane (1984) RCT	Inpatient hospice plus home care vs conventional hospital care	Single institution	Hospital, home care (insurance claims)	Pain, depression, anxiety, satisfaction, functioning	No significant differences in cost or effectiveness with significantly higher satisfaction ratings for home care.
4	Rafferty (1996) Observational study	Coordination service by nurses to arrange care vs conventional care	Social costs. (but excluding out-of-pocket expenses)	Hospital, home care, GP, social services,	Depression, physical symptoms, satisfaction,	No measurable difference in outcome, fewer hospital days (crude estimates only). Costs not reported alongside outcomes.
5	Tramartin (1992) Quasi randomised study <sup>2</sup>	Home care vs hospital care	Single institution	Average cost for the specific programme only (no data on individuals)	Modified utility measure: Quality of Well-Being	Similar QWB scores. Estimated cost savings of 34% for home care
6	Hughes (1992) RCT	Home care vs hospital and community care	Multiple sectors	Hospital, home care, nursing home	Self care, mental health, function, ADL, morale, satisfaction	Increased home care costs were offset by reduction in hospitalisation. Significantly higher satisfaction reported for the intervention group. No difference in morale.
7	Bloom (1980) Prospective case control	Home care vs hospital care	Multiple sectors	Hospital, home-maker services, loss of income in home care group only	Qualitative evidence from carers	Total cost significantly lower for intervention group for the two-week period before death for those who died at home (versus hospital). Costs only reported.
8/9	Zimmer (1984) RCT	Physician-led home care vs usual health and home care	Multiple sectors	Hospital, home care, "related services"	Sickness impact profile, morale, place of death, carer satisfaction	Trend (non-significant) towards lower hospitalisation explained by lower costs of patients who died during the study. Only score that was significantly higher for home care was for satisfaction.
10	McCusker (1987) Retrospective case control	Home care vs usual care	Single institution	Hospital and home care only	Resource utilisation only	Savings among users achieved by decreasing hospital days and reduction in mean daily cost of hospitalisation. Greatest benefits for those close to time of death. Costs not reported alongside outcomes.
11	Axelsson (1998) Before and after study	Inpatient palliative support team conventional care	Society	Hospital, home care, district nursing, GP, social services, night sitting, loss of income	Net saving in institutional days. No quality of life assessment	Reduction in both hospitalisation and home care days. Costs only reported

<sup>1</sup> Activities of daily living

<sup>2</sup> Only the first 10 patients randomised by number and who met selection criteria were offered home care

12	Ventafriidda (1989 ) Prospective case control	Home care vs hospital palliative care	Multiple sector	Hospital, home care, volunteers, private nursing	Pain and number of symptoms, ADL, satisfaction	Higher levels of quality of life reported after two weeks on pain and symptom scale, morale, and satisfaction. Home care was also less expensive.
13	Greer (1986) Observational study	Inpatient hospice and home care vs conventional oncology care in hospital	Single institution	Hospital and home care	ADL, symptom severity, satisfaction, % analgesia prescribed, carer morbidity, anxiety, depression and emotional distress post-bereavement	Few robust quality of life advantages associated with hospice. Hospital based hospice may reduce costs, hospice home care does not. The longer the patients are in hospice the fewer savings accrue. Savings were due to patients who died incurring lower costs in hospice. Conventional care patients were more likely to receive intensive treatments before death, have more pain and symptoms. No difference in satisfaction. More carer satisfaction with hospice.
14	Dunt (1989) Quasi experimental case control	Inpatient hospice and home care vs specialist home care vs general nursing vs hospital oncology ward	Multiple sector	All health care use	Pain, ADL, Symptoms, quality of life and satisfaction.	Significantly lower pain duration for hospice patients, higher satisfaction with care, lower carer burden, no differences in quality of life, Costs not assessed alongside outcomes. No savings demonstrated due to high labour and administrative costs for home hospice care
15	McCorckle (1989) Longitudinal experimental study	Specialist and non-specialist home care vs physician office care only	Single institution	Hospital	Symptoms, distress, pain, inventory of current concerns, Mood profile, social dependency, general health, medical record review	No significant difference in pain, current concerns, or mood profiles. Symptom distress similar for both home groups, , remaining independent for longer, fewer hospital days for specialist home care patients (non-significant)

## Appendix B

Copies of patients information sheets and consent forms used in the study

### [Letter to patient from the hospice]

Dear

I am writing to ask for your help in a research project we are doing. The project is to find out whether patients benefit from going to day care centres or whether they receive just as good services at home and in hospital. We hope to interview lots of patients who use different kinds of services to find out whether day care makes a difference to patients' health or not.

We are contacting patients who use day care centres and patients who do not so we can see whether day care makes a difference. If you would be willing to take part in the study, please fill in the attached form and send it to the address written at the bottom of the form. Then one of the trained interviewers will contact you and make an arrangement to come and explain the study to you more fully. You will have a chance then to ask any questions you want and the interviewing will not start before you are happy that you understand what will happen and have agreed to take part.

I am delighted that the Department of Health has funded this study so that we can ask patients and their carers what they think about the health services they receive and how it can be improved for patients in the future. I hope you will be able to take part and give your views.

With best wishes

Yours sincerely

Palliative Care Consultant

**Patient information sheet given to patients by their nurse prior to interview**

I am a researcher working for Guy, King's and St Thomas' Medical School in South London\*. I would like to invite you to take part in a study, funded by The NHS to evaluate palliative [day\*] care.

If you agree to take part in the study I will want to know about the care that you will receive. I will firstly ask you to complete an interview with me which consists of three questionnaires that ask you about how you are feeling, and another to tell me about the health care you have been receiving. This first visit should last for half an hour. I would then like to see you two more times, once in about 6-8 weeks, once in about three months time. I will contact you prior to each visit to make sure you are still happy to be interviewed. At each interview I will ask you to complete the same short questionnaires. I will also ask you if you think you have changed over time, what is good about the care you receive, and what you would like to change. These visits should only take 20 minutes to half an hour.

Your Macmillan nurse / palliative care home nurse\* will have told you a bit about the study and asked if you would be willing to see me. Any other professionals who may be involved in your care does not have to know you have seen me, if this is what you would like.

All the information collected in this study will be strictly confidential. Your participation in this study is entirely voluntary and if you do take part, you are free to withdraw at any time. Your involvement or non involvement in this study will not influence the care you receive. If you are happy to be involved, but do not wish us to contact your carer or any other professional about your care, then we will not do so.

It is hoped that the information from this study will help us gain knowledge about palliative care and how best to meet your needs. I will come to see you soon to see if you are interested in taking part. If you or your family have any further questions please feel free to contact me on the number below or ask the ward nurse / home care nurse to contact me

Thank you for considering taking part in our research.

Yours faithfully etc.

*\* alter as needed*



## **Patient information sheet for the choice experiment study**

### **What is the study for?**

We are undertaking a research project to assess the needs to patients who use the palliative day care services here at XXXXX. This is part of a study involving four similar centres in London and the South-East.

The purpose of this study is to find out about why people come to palliative day care and what they want from the service. Day services are not the same in all centres. We want to find out how they should be organised to suit your needs. We do not need to know anything about your illness and anything you say to the interviewer will be treated in the strictest confidence.

The study is funded by the NHS. The researcher does not work at XXXX but is a trained interviewer with palliative care care nursing experience

### **The questionnaire**

On each sheet of the questionnaire there are simple descriptions of two different palliative day therapy centres. They are open at different hours and offer different kinds of services. We would like you to choose which one you would attend if you had the choice. Each set of pairs offers a slightly different choice. From this, we can find out how strong your views are about the services you want.

### **What do you have to do?**

First, you need to sign the consent form overleaf to say you are happy to take part in the study.

The descriptions will be read out to you, or you can read it through yourself, whichever you are most comfortable with. Take your time to consider the options and think about which one you would choose to attend. Tick the box at the bottom of each page indicating your choice, A or B.

There are also some simple questions about who you are and how often you come to XXXX that we would like you to fill in.

### **If you have any problems....**

Let the interviewer know that you'd like a break whenever you feel you need one. You don't have to finish and you can stop whenever you want, for whatever reason.

***Thank you.***

**Consent form for the palliative day care study**

I have read or been told about the research project I have been asked to participate in and have had all my questions answered. I understand the purpose of this study and have had questions answered to my satisfaction. I understand that my name and address will not be identified in the research and everything I say will be treated in the strictest confidence.

[Choice experiment study only] I will not be asked any personal questions about my illness.

Nothing I say to the researcher will be passed on to the staff of the centre, unless I give my consent for it to do so. All the data will be kept by the researcher and not given to anyone else. I can stop the interview whenever I feel like.

Signature \_\_\_\_\_

Name in print: \_\_\_\_\_

[Witnessed by\* ] \_\_\_\_\_

Date \_\_\_\_\_

*\* where respondent unable to write own signature and has agreed verbally to take part in the study*

Appendix C Copies of research instruments

**Semi-structured interview for patients on use of services**

I'd like to begin by asking you to tell me about the frequency and types of services you have received over the *last month* not including day care. (*Let me explain what I mean by day care – your visit to the hospice day centre and the activities that you do while you are there*).

So over the *last month*, whom have you seen or spoken to from health or social services, and *how often*? (*Please note the frequency and type of contact*).

<b>Nurses:</b>		
A district (or community) nurse:	Yes	No
A nurse at/from the general practice:	Yes	No
A specialist palliative care nurse at home: (Macmillan or hospice nurse)	Yes	No
A specialist palliative care nurse at home: (E.g. Marie-Curie nurse)	Yes	No
A specialist palliative care nurse in hospital:	Yes	No
Any other nurse (specify): (Remember not from day care)	Yes	No
<b>Doctors:</b>		
Your GP:	Yes	No
Specialist doctor <u>at home</u> :	Yes	No

**(From a hospice, a hospital or palliative care team)**

<b>Specialist Hospital doctor:</b>	Yes	No
<b>Other doctors (specify):</b> (Remember not from day care)	Yes	No

**Other people: Who else have you seen or spoken to over the last month?**  
(Not from day care)

<b>Occupational therapist:</b>	Yes	No
<b>Health visitor:</b>	Yes	No
<b>Dietician:</b>	Yes	No
<b>Physiotherapist:</b>	Yes	No
<b>Other:</b> (Specify type & frequency)	Yes	No

Further notes:

*Please specify other social services – e.g. MOW, H/H or family/social network.*

Over the *last month*, how many nights have you spent away from home for health-related reasons:

In a hospital:	_____
In a hospice:	_____
Elsewhere (specify):	_____

Over the *last month*, how many times have you visited as an outpatient in addition to the day care you are currently receiving:

The hospital:	_____
The hospice:	_____
Day care:	_____
Other (specify):	_____

**Are there any services that you don't receive now that you used to, or receive less of?**

**Are there any services you now receive more of?**

<b>POS Questionnaire</b>
--------------------------

Answer the following questions by circling the answer, which you think most accurately, describes how you have been feeling.

**1. Over the past 3 days, have you been affected by pain?**

Not at all, no effect	0
Slightly - but not bothered to be rid of it	1
Moderately - pain limits some activity	2
Severely - activities or concentration markedly affected	3
Overwhelmingly - unable to think of anything else	4

**2. Over the past 3 days, have other symptoms e.g. feeling sick, having a cough or constipation been affecting how you feel?**

No, not at all	0
Slightly	1
Moderately	2
Severely	3
Overwhelmingly	4

**3. Over the past 3 days, have you been feeling anxious or worried about your illness or treatment?**

No, not at all	0
Occasionally	1
Sometimes - affects my concentration now and then	2
Most of the time - often affects my concentration	3
Can't think of anything else - completely preoccupied	4

**4. Over the past 3 days, have any of your family or friends been anxious or worried about you?**

No, not at all	0
Occasionally	1
Sometimes - it seems to affect their concentration	2
Most of the time	3
Yes, always preoccupied with worry about me	4

**5. Over the past 3 days, how much information have you and your family or friends been given?**

Full information - always feel free to ask what I want	0
Information given but hard to understand	1

Information given on request but would have liked more	2
Very little given and some questions were avoided	3
None at all	4

**6. Over the past 3 days, have you been able to share how you are feeling with your family or friends?**

Yes, as much as I wanted to	0
Most of the time	1
Sometimes	2
Occasionally	3
No, not at all with anyone	4

**7. Over the past 3 days, have you felt that life was worthwhile?**

Yes, all the time	0
Most of the time	1
Sometimes	2
Occasionally	3
No, not at all	4

**8. Over the past 3 days, have you felt good about yourself as a person?**

Yes, all the time	0
Most of the time	1
Sometimes	2
Occasionally	3
No, not at all	4

**9. Over the past 3 days, how much time do you feel you have wasted on appointments relating to your healthcare, e.g. waiting around for transport or having the same tests repeated?**

None at all	0
Up to half a day wasted	1
More than half a day wasted	2

**10. Over the past 3 days, have any practical matters resulting from your illness, either financial or personal, been addressed?**

Practical problems have been addressed and my affairs are as up to date as I would wish	0
Practical problems are in the process of being addressed	1
Practical problems exist which were not addressed	2
I have had no practical problems	3

**11. If any, what have been your main problems in the last three days?**

Herth Hope Index

	Strongly disagree	Disagree	Agree	Strongly agree
In have a positive outlook toward life				
I have short and/or long range goals				
I feel all alone				
I can see possibilities in the midst of difficulties				
I have faith that gives me comfort				
I feel scared about my future				
I can recall happy/joyful times				
I have deep inner strength				
I am able to give and receive caring love				
I have a sense of direction				
I believe that each day has potential				
I feel my life has value and worth				



**Qualitative questions**

Can you tell me what is the best thing about attending this centre?

Do you think you have changed as a result of attending palliative day care?

Yes/ No

Can you explain why

Do you think there any down sides to attending palliative day care?

**EQ-5D instrument**

By placing a tick in one box below, please indicate which statements best describe your own health state today

**Mobility**

- I have no problems in walking about ☐
- I have some problems in walking about ☐
- I am confined to bed ☐

**Self-care**

- I have no problems with self-care ☐
- I have some problems washing or dressing myself ☐
- I am unable to wash or dress myself ☐

**Usual activities** (*e.g. work, study, housework, family or leisure activities*)

- I have no problems with performing my usual activities ☐
- I have some problems with performing my usual activities ☐
- I am unable to perform my usual activities ☐

**Pain/discomfort**

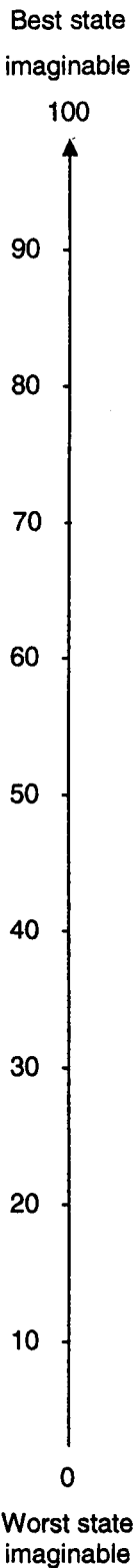
- I have no pain or discomfort ☐
- I have moderate pain or discomfort ☐
- I have extreme pain or discomfort ☐

**Anxiety/ depression**

- I am not anxious or depressed ☐
- I am moderately anxious or depressed ☐
- I am extremely anxious or depressed ☐

To help people say how good or bad a health state is, we have drawn a scale (rather like a thermometer) on which the best state you can imagine is marked 100 and the worst state you can imagine is marked 0. We would like you to indicate on this scale how good or bad your own health is today, in your opinion. Please do this by drawing a line from the box below to whichever point on the scale indicates how good or bad your health is today.

Your own  
health state  
today



**Appendix D Data collection form for the choice experiment study**

Patient name: last name) \_\_\_\_\_(first name) \_\_\_\_\_  
Date: \_\_\_\_\_

Date of birth: \_\_\_\_\_ Age \_\_\_\_\_  
(day/month/year)

Date of first visit to centre (approximately): \_\_\_\_\_  
(month/year)

Does the patient live alone? Yes ☐ No ☐

If not, do they live with:

- a. spouse
- b. family member
- c. other \_\_\_\_\_

How often does the patient attend Day Care Centre?

- a. Weekly (which day of the week? \_\_\_\_\_)
- b. 2 times/ week (on \_\_\_\_\_ & \_\_\_\_\_)
- c. 3 times/ week (on \_\_\_\_\_, \_\_\_\_\_, & \_\_\_\_\_)
- d. other: \_\_\_\_\_

When the patient comes to Day Care, how long do they usually stay?

- a. All day (10 to 3)
- b. a.m. only
- c. p.m. only
- d. For appt. only

The patient participates in which of the following activities/services? (circle all that apply)

- |                |               |                     |
|----------------|---------------|---------------------|
| See the doctor | Counselling   | Social interactions |
| Arts & crafts  | Music Therapy | Aromatherapy        |
| Music Therapy  | Massage       |                     |
| Physiotherapy  | Bathing       |                     |
| Hairdressing   | Reflexology   |                     |

Is there anything special the researcher needs to know about this patient, for example, state of health or special health concerns? (carry on overleaf if necessary)

**Choice experiment interview schedule**

“Imagine the two palliative day care centres you could attend once a week. One is centre A and one is centre B. Both offer arts and crafts and a friendly, social environment. But other than that they are not the same and offer different kinds of services and activities.

Some of the choices may appear easy; others will be more difficult. Just tell us which, if you had to choose, you would prefer to visit?”

Card 1

	Centre A	Centre B
Access	You can only come for an hour or so but not stay all day	Stay for the full session
Time	10 am to 3 pm	1 pm to 6 pm
Bathing	You can have a bath every visit if you want to	Not available
Specialised therapies and activities	Not available	Massage, aromatherapy and reflexology are available
Hairdressing	Not available	Hairdressing is available
Doctor	In emergencies only	You can see a doctor every visit if you want to

Which centre do you prefer?

A

B

☐☐

These choices are slightly different, which would you choose now?

Card 2

	Centre A	Centre B
Access	You can only come for an hour or so but not stay all day	Stay for the full session
Time	10 am to 3 pm	1 pm to 6 pm
Bathing	Not available	You can have a bath every visit if you want to
Specialised therapies and activities	Massage, aromatherapy and reflexology are available	Not available
Hairdressing	Hairdressing is available	Not available
You can see a doctor every visit if you want	In emergencies only	You can see a doctor every visit if you want to

Which centre do you prefer?

A

B

☐☐

Card 3

	Centre A	Centre B
Access	Stay for the full session	You can only come for an hour or so but not stay all day
Time	10 am to 3 pm	1 pm to 6 pm
Bathing	Not available	You can have a bath every visit if you want to
Specialised therapies and activities	Not available	Massage, aromatherapy and reflexology are available
Hairdressing	Hairdressing is available	Not available
Doctor	You can see a doctor every visit if you want to	In emergencies only

Which centre do you prefer?

A

B

☐☐

Card 4

	Centre A	Centre B
Access	Stay for the full session	You can only come for an hour or so but not stay all day
Time	10 am to 3 pm	1 pm to 6 pm
Bathing	You can have a bath every visit if you want to	Not available
Specialised therapies and activities	Massage, aromatherapy and reflexology are available	Not available
Hairdressing	Not available	Hairdressing is available
You can see a doctor every visit if you want	You can see a doctor every visit if you want to	In emergencies only

Which centre do you prefer?

A

B

☐☐



Card 5

	Centre A	Centre B
Access	Stay for the full session	You can only come for an hour or so but not stay all day
Time	1 pm to 6 pm	10 am to 3 pm
Bathing	You can have a bath every visit if you want to	Not available
Specialised therapies and activities	Not available	Massage, aromatherapy and reflexology are available
Hairdressing	Hairdressing is available	Not available
Doctor	In emergencies only	You can see a doctor every visit if you want to

Which centre do you prefer?

A

B

☐☐

**STOP HERE AND TAKE A BREAK**

Card 6

	Centre A	Centre B
Access	Stay for the full session	You can only come for an hour or so but not stay all day
Time	1 pm to 6 pm	10 am to 3 pm
Bathing	Not available	You can have a bath every visit if you want to
Specialised therapies and activities	Massage, aromatherapy and reflexology are available	Not available
Hairdressing	Not available	Hairdressing is available
Doctor	In emergencies only	You can see a doctor every visit if you want to

Which centre do you prefer?

A

B

☐☐

Card 7

	Centre A	Centre B
Access	You can only come for an hour or so but not stay all day	Stay for the full session
Time	1 pm to 6 pm	10 am to 3 pm
Bathing	You can have a bath every visit if you want to	Not available
Specialised therapies and activities	Massage, aromatherapy and reflexology are available	Not available
Hairdressing	Hairdressing is available	Not available
Doctor	You can see a doctor every visit if you want to	In emergencies only

Which centre do you prefer?

A

☐

B

☐

Card 8

	Centre A	Centre B
Access	You can only come for an hour or so but not stay all day	Stay for the full session
Time	1 pm to 6 pm	10 am to 3 pm
Bathing	Not available	You can have a bath every visit if you want to
Specialised therapies and activities	Not available	Massage, aromatherapy and reflexology are available
Hairdressing	Not available	Hairdressing is available
Doctor	You can see a doctor every visit if you want to	In emergencies only

Which centre do you prefer?

A

B

☐☐

Card 9

	Centre A	Centre B
Access	You can only come for an hour or so but not stay all day	Stay for the full session
Time	10 am to 3 pm	1 pm to 6 pm
Bathing	Not available	You can have a bath every visit if you want to
Specialised therapies and activities	Not available	Massage, aromatherapy and reflexology are available
Hairdressing	Hairdressing is available	Not available
Doctor	You can see a doctor every visit if you want to	In emergencies only

Which centre do you prefer?

A

☐

B

☐

Card 10

	Centre A	Centre B
Access	You can only come for an hour or so but not stay all day	Stay for the full session
Time	10 am to 3 pm	1 pm to 6 pm
Bathing	You can have a bath every visit if you want to	Not available
Specialised therapies and activities	Massage, aromatherapy and reflexology are available	Not available
Hairdressing	Not available	Hairdressing is available
Doctor	In emergencies only	You can see a doctor every visit if you want to

Which centre do you prefer?

A

☐

B

☐

## Appendix E Output of the STATA and SPSS anlyiss of the choice experiment probit and logit models

### A) STATA analysis

```
. probit choicel diffacc diffbat diffmass diffhairdr diffdr difftime if scenario <9
```

```
Iteration 0: log likelihood = -432.47256
Iteration 1: log likelihood = -345.80258
Iteration 2: log likelihood = -343.40017
Iteration 3: log likelihood = -343.38204
Iteration 4: log likelihood = -343.38203
```

Probit estimates	Number of obs	=	624
	LR chi2(6)	=	178.18
	Prob > chi2	=	0.0000
Log likelihood = -343.38203	Pseudo R2	=	0.2060

choicel	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
diffacc	.2831787	.0567376	4.99	0.000	.1719751	.3943824
diffbat	.0811388	.0568552	1.43	0.154	-.0302954	.192573
diffmass	.6117827	.0568678	10.76	0.000	.5003239	.7232415
diffhairdr	.0445646	.0564098	0.79	0.430	-.0659965	.1551258
diffdr	.185606	.0561943	3.30	0.001	.0754672	.2957448
difftime	-.2987058	.056885	-5.25	0.000	-.4101984	-.1872132
_cons	.0220469	.0561231	0.39	0.694	-.0879523	.1320461

```
. xtprobit choicel diffacc diffbat diffmass diffhairdr diffdr difftime if scenario <9,
i(pt_code)
```

Fitting comparison model:

```
Iteration 0: log likelihood = -432.47256
Iteration 1: log likelihood = -345.80258
Iteration 2: log likelihood = -343.40017
Iteration 3: log likelihood = -343.38204
Iteration 4: log likelihood = -343.38203
```

Fitting full model:

```
rho = 0.0 log likelihood = -343.38203
rho = 0.1 log likelihood = -350.35116
Iteration 0: log likelihood = -343.38209
Iteration 1: log likelihood = -343.38209
Iteration 2: log likelihood = -343.38209
```

Random-effects probit	Number of obs	=	624
Group variable (i) : pt_code	Number of groups	=	79

Random effects u_i ~ Gaussian	Obs per group: min	=	4
	avg	=	7.9
	max	=	8

Log likelihood = -343.38209	Wald chi2(6)	=	147.08
	Prob > chi2	=	0.0000

choicel	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
diffacc	.2831787	.0567376	4.99	0.000	.1719751	.3943824
diffbat	.0811388	.0568552	1.43	0.154	-.0302954	.192573
diffmass	.6117827	.0568678	10.76	0.000	.5003239	.7232415
diffhairdr	.0445646	.0564098	0.79	0.430	-.0659965	.1551258
diffdr	.185606	.0561943	3.30	0.001	.0754672	.2957448
difftime	-.2987058	.056885	-5.25	0.000	-.4101984	-.1872132
_cons	.0220469	.0561232	0.39	0.694	-.0879525	.1320463
/lnsig2u	-14	140.8855			-290.1305	262.1305

sigma_u		.0009119	.0642355		9.98e-64	8.34e+56
rho		8.32e-07	.0001172		1.0e-126	1

-----  
Likelihood ratio test of rho=0: chibar2(01) = 0.00 Prob >= chibar2 = 1.000

end of do-file

. xi:probit choicel diffacc diffbat diffmass diffhairdr diffdr difftime if scenario <9

Iteration 0: log likelihood = -432.47256  
Iteration 1: log likelihood = -345.80258  
Iteration 2: log likelihood = -343.40017  
Iteration 3: log likelihood = -343.38204  
Iteration 4: log likelihood = -343.38203

Probit estimates	Number of obs	=	624
	LR chi2(6)	=	178.18
	Prob > chi2	=	0.0000
Log likelihood = -343.38203	Pseudo R2	=	0.2060

choicel	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
diffacc	.2831787	.0567376	4.99	0.000	.1719751 .3943824
diffbat	.0811388	.0568552	1.43	0.154	-.0302954 .192573
diffmass	.6117827	.0568678	10.76	0.000	.5003239 .7232415
diffhairdr	.0445646	.0564098	0.79	0.430	-.0659965 .1551258
diffdr	.185606	.0561943	3.30	0.001	.0754672 .2957448
difftime	-.2987058	.056885	-5.25	0.000	-.4101984 -.1872132
_cons	.0220469	.0561231	0.39	0.694	-.0879523 .1320461

. logit choicel diffacc diffbat diffmass diffhairdr diffdr difftime if scenario <9

Iteration 0: log likelihood = -432.47256  
Iteration 1: log likelihood = -346.49075  
Iteration 2: log likelihood = -343.21072  
Iteration 3: log likelihood = -343.09296  
Iteration 4: log likelihood = -343.09265

Logit estimates	Number of obs	=	624
	LR chi2(6)	=	178.76
	Prob > chi2	=	0.0000
Log likelihood = -343.09265	Pseudo R2	=	0.2067

choicel	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
diffacc	.4953618	.1000917	4.95	0.000	.2991856 .691538
diffbat	.1655054	.1008856	1.64	0.101	-.0322267 .3632375
diffmass	1.027025	.1008236	10.19	0.000	.8294143 1.224635
diffhairdr	.051938	.0984523	0.53	0.598	-.1410249 .2449009
diffdr	.3235665	.0971495	3.33	0.001	.133157 .513976
difftime	-.5184057	.1011059	-5.13	0.000	-.7165696 -.3202418
_cons	.0521439	.0970193	0.54	0.591	-.1380104 .2422982



```
. xtlogit choicel diffacc diffbat diffmass diffhairdr diffdr difftime if scenario <9,
i(pt_code)
```

Fitting comparison model:

```
Iteration 0: log likelihood = -432.47256
Iteration 1: log likelihood = -346.49075
Iteration 2: log likelihood = -343.21072
Iteration 3: log likelihood = -343.09296
Iteration 4: log likelihood = -343.09265
```

Fitting full model:

```
rho = 0.0 log likelihood = -343.09265
rho = 0.1 log likelihood = -345.62132
Iteration 0: log likelihood = -343.09267
Iteration 1: log likelihood = -343.09267
```

```
Random-effects logit                                Number of obs      =      624
Group variable (i) : pt_code                        Number of groups   =       79

Random effects u_i ~ Gaussian                        Obs per group: min =        4
                                                    avg =       7.9
                                                    max =        8

Log likelihood = -343.09267                          Wald chi2(6)       =    123.08
                                                    Prob > chi2        =    0.0000
```

choicel	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
diffacc	.4953618	.100092	4.95	0.000	.299185	.6915386
diffbat	.1655054	.1008859	1.64	0.101	-.0322273	.3632381
diffmass	1.027025	.1008239	10.19	0.000	.8294136	1.224636
diffhairdr	.051938	.0984525	0.53	0.598	-.1410253	.2449013
diffdr	.3235665	.0971497	3.33	0.001	.1331566	.5139764
difftime	-.5184057	.1011062	-5.13	0.000	-.7165703	-.3202411
_cons	.0521439	.0970195	0.54	0.591	-.1380109	.2422987
/lnsig2u	-14	231.2352			-467.2126	439.2126
sigma_u	.0009119	.1054296			3.5e-102	2.36e+95
rho	2.53e-07	.0000178			3.8e-204	1

Likelihood ratio test of rho=0: chibar2(01) = 0.00 Prob >= chibar2 = 1.000

```
. xi:logit choicel diffacc diffbat diffmass diffhairdr diffdr difftime if scenario <9
```

```
Iteration 0: log likelihood = -432.47256
Iteration 1: log likelihood = -346.49075
Iteration 2: log likelihood = -343.21072
Iteration 3: log likelihood = -343.09296
Iteration 4: log likelihood = -343.09265
```

```
Logit estimates                                Number of obs      =      624
                                                    LR chi2(6)        =    178.76
                                                    Prob > chi2       =    0.0000
Log likelihood = -343.09265                    Pseudo R2         =    0.2067
```

choicel	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
diffacc	.4953618	.1000917	4.95	0.000	.2991856	.691538
diffbat	.1655054	.1008856	1.64	0.101	-.0322267	.3632375
diffmass	1.027025	.1008236	10.19	0.000	.8294143	1.224635
diffhairdr	.051938	.0984523	0.53	0.598	-.1410249	.2449009
diffdr	.3235665	.0971495	3.33	0.001	.133157	.513976
difftime	-.5184057	.1011059	-5.13	0.000	-.7165696	-.3202418
_cons	.0521439	.0970193	0.54	0.591	-.1380104	.2422982

end of do-file

**B) SPSS analysis**

## DATA Information

624 unweighted cases accepted.

0 cases rejected because of missing data.

0 cases are in the control group.

## MODEL Information

ONLY Normal Sigmoid is requested.

----- \* \*  
 \* \* \* \* \* P R O B I T     A N A L Y S I S \* \* \* \* \*

Parameter estimates converged after 15 iterations.  
 Optimal solution found.

Parameter Estimates (PROBIT model: (PROBIT(p)) = Intercept + BX):

	Regression Coeff.	Standard Error	Coeff./S.E.
DIFFACC	.28318	.05674	4.99103
DIFFBAT	.08114	.05686	1.42712
DIFFMASS	.61178	.05687	10.75795
DIFFHAIR	.04457	.05641	.79004
DIFFDR	.18560	.05619	3.30286
DIFFTIME	-.29871	.05688	-5.25106

Intercept	Standard Error	Intercept/S.E.
.02204	.05612	.39277

Pearson Goodness-of-Fit Chi Square = 632.339    DF = 617    P = .326

Since Goodness-of-Fit Chi square is NOT significant, no heterogeneity factor is used in the calculation of confidence limits.